Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

Bureau of Educational and Psychological
Govt. of W. Bengal
Library
Calcutta

VOL. 73
1959

NORMAN L. MUNN, EDITOR
Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
FRANCIS W. IRWIN
JAMES J. JENKINS
HAROLD E. JONES
DANIEL KATZ
BOYD McCANDLESS
DONALD W. MACKINNON
QUINN McNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
RICHARD L. SOLOMON
ROSS STAGNER

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Bureau of Edn! & Psyl. Research
(S. C. E. R. T.)
Date 9. 3. 81
Acc. J739

CONTENTS OF VOLUME 73

Whole No.

471 Mass Media and Children: A Study of Exposure Habits and Cognitive Effects. Lotte Bailyn. Pp. 48.

472 The Effects of Early Experience on Subsequent Emotionality and Resistance to Stress. Robert Ader. Pp. 31.

473 Quasi-Therapeutic Relations in Small College and Military Groups. Fred E. Fiedler, Edwin B. Hutchins, and Joan S. Dodge. Pp. 28.

474 The Influence of Different Activities on the Apparent Length of Time. John C. Loehlin. Pp. 27.

475 The Ontogeny of Certain Logical Operations: Piaget's Formulation Examined by Nonverbal Methods. Martin D. S. Braine. Pp. 43.

476 Congruencies Among Interpretations of Psychological Test and Anamnestic Data. Kenneth B. Little and Edwin S. Shneidman. Pp. 42.

477 A Q-Sort Study of the Validity of Evaluations Made From Projective Techniques. Lloyd H. Silverman. Pp. 28.

478 The Influence of Classification Requirements on Gradients of Response. Michael A. Wallach. Pp. 21.

479 A Factor-Analytic Study Across the Domains of Reasoning, Creativity, and Evaluation. Norman W. Kettner, J. P. Guilford, and Paul R. Christensen. Pp. 31.

480 Group Psychotherapy and Concomitant Psychological Change. Earl J. Ends and Curtis W. Page. Pp. 31.

481 Photosensitization in Relation to Mean and Standard Deviation Values. Richard T. Zegers. Pp. 25.

482 Correlates of Later Performance and Specialization in Psychology: A Follow-up Study of the Trainees Assessed in the VA Selection Research Project. E. Lowell Kelly and Lewis R. Goldberg. Pp. 32.

483 Personality, Intellectual, and Achievement Patterns in Gifted Children. Adma d'Heurle, Jeanne Cummins Mellinger, and Ernest A. Haggard. Pp. 28.

484 Impaired Performance with Acute Sleep Loss. Harold L. Williams, Ardie Lubin, and Jacqueline J. Goodnow. Pp. 26.

485 Personality and Environmental Factors in the Development of Delinquency. Robert D. Wirt and Peter F. Briggs. Pp. 47.

486 Imitative Behavior in Preschool Children. John W. McDavid. Pp. 26.

487 The Validity of the Thematic Apperception Test in the Study of Adolescent Personality. William E. Henry and Jane Farley. Pp. 40.

Bureau Ednl. Psy. Res
DAVID HA[illegible] AINING COL[illegible]GE
Dated
No.

Vol. 73, No. 1 Whole No. 471, 1959
Psychological Monographs: General and Applied

MASS MEDIA AND CHILDREN:
A STUDY OF EXPOSURE HABITS AND COGNITIVE EFFECTS[1]

LOTTE BAILYN
Department of Social Relations, Harvard University

> **And shall we just carelessly allow children to hear any casual tales which may be devised by casual persons, and to receive in their minds ideas for the most part the very opposite of those which we should wish them to have when they are grown up?**
>
> —PLATO, *The Republic*, Book II

INTRODUCTION

THE PROBLEM of the effect on children of such "casual tales" has assumed great importance in our time with the advent of mass media of communication. Over the past three decades, concern about this problem has been reflected in passionately expressed opinions about the good and evil of every new mass medium, as well as in the more dispassionate efforts of research workers. The latter, however, have found the problem of effects an elusive one because of the variety and complexity of the factors involved.

Meaning of "Effects"

The area of effects is a broad one, and any investigation of it must clarify the limits within which to proceed. It should be noted, in the first place, that effects of the mass media on children can be both general and specific. In speaking of general effects one is concerned with the psychological and social problems a child faces in a society in which the media play an important role. Specific effects stem from the personal exposure of a particular child. The present discussion is limited to the latter. It should be recognized, secondly, that effects may be direct or indirect: the direct having to do with effects of the mass media on characteristics within the child himself and the indirect with influences on others whose behavior affects him. The present discussion centers on the former.

Speaking, then, of direct and specific effects, one must consider, further, the area within which such effects are to be sought. Of the possible alternatives the most socially relevant area is, no doubt, that of overt behavior. Nevertheless, though the understanding of this connection may be the eventual goal, it is most fruitful at present to concentrate on some of the many mediating factors between exposure to material in the mass media and the translation of its influence into overt action. Among these mediating elements are certain of the child's habits of thought, and it is with the effects of the mass media on these habits that this study is centrally concerned.

Use of the word "effects" to describe the problem of this research does not imply the assumption of a simple causal connection between the act of exposure and a way of thinking that correlates positively with it. Such a correlation, rather, is assumed to be the end result of a process consisting of increments in exposure interacting with increments in the particular way of thinking. For a cognitive correlate—a way of thinking, positively correlated with exposure—to be considered an "effect," one should have evidence that an increment in exposure increases the tendency to think in the given way. With such evidence one could reject the hypothesis that the original correlation results from self-selection alone—from the fact that children with a particular cognitive bent expose themselves more. Such evidence would suggest that at least part of any such correlation involves a causal link starting with exposure and ending with a particular way of thinking. It would suggest, in other words, that the particular character of the thought process would be absent but for material in the mass media. The present study deals with effects in this sense.

---

[1] This monograph is a revised version of a doctoral dissertation written under the direction of Gordon W. Allport of Harvard University. It was partially supported by funds from a grant of the Council of Mental Health of the American Medical Association to Boston University made available to me by Nathan Maccoby of Boston University.

I am deeply grateful to Professor Allport for his wise counsel, his patient help, and his encouragement in the planning and writing of the dissertation and the subsequent revisions. I had the benefit also of the guidance of Eleanor E. Maccoby of Harvard University. She allowed me to lean heavily on her experience and skill in research design and analysis; her help throughout was of inestimable value. It is a pleasure also to acknowledge the generous cooperation of Bert A. Roens, Assistant Superintendent of the Arlington Massachusetts Public Schools, and of the principals, teachers, and pupils of those schools. Their cooperation made the study possible.

Specific Problem

The thought processes that might be studied in such a context are numerous. The present selection is based partly on concerns expressed about children's exposure to the mass media and partly on an analysis of elements that occur frequently in mass media content for children.

The final selection consists of four aspects of thinking. Two of these reflect fairly directly specific elements of certain stories frequently found in the mass media. The first is the tendency to stereotype. Many comic books and television programs for children present personality crudely and simply. A person is either good or bad. Which category he falls into is determined more by fiat and conventional expression than by psychologically meaningful motivations. Do children who are constantly exposed to such stories pick up this way of classifying people? Does this feature of the content affect the degree of stereotyping in their thinking?

The second aspect, closely related to content, concerns the perception by the child of the existence of a general threat to him as an individual or to his country or to the world. Many stories present the individual as helpless in this world. Even the law-enforcement officer frequently is powerless in the face of evil and has to be helped by the supermen of children's stories. Does this element of content carry over to the child's perception of the power of the individual to control his own destiny?

The two other aspects dealt with here have to do with the influence of exposure to the mass media on (*a*) a child's self-image and (*b*) the degree of passivity in his outlook. Specifically, the former relates to the following question: If one asks a child what he wants to be when he is grown up or whom he would most like to be changed into, does he draw on the personalities frequently portrayed in the mass media for his answers? Two related questions are asked about the problem of passivity: first, concerning the preference for types of activities of a more or less passive nature; second, concerning the passiveness implicit in desiring an occupation the same as, or similar to, that of the parents.

The central problem of this study, then, is the effect of the mass media on four cognitive aspects of the child: stereotyping, perception of threat, projected self-image, and passivity.

General Procedure

The amount of a child's exposure and the material to which he is exposed are necessary data for any study of effects. Also important is the child's perception of this content, hence this, too, must be taken into account if effects are to be understood.

In addition to the child's media habits, certain other characteristics must be considered. Obviously, the way a child thinks is to a large extent dependent on his established psychological characteristics and on the social characteristics of the groups to which he belongs as well as on the mass media. Ignoring such influences on thought could result in the establishment of spurious relationships, or it could indicate mistakenly a lack of relationship between exposure to the mass media and thought processes. Therefore a questionnaire[2] was administered to over six hundred boys and girls in the fifth and sixth grades of the public schools of a suburb of Boston, Massachusetts, that yielded basic information on exposure habits, social background, and psychological characteristics as well as evidence relevant to the cognitive aspects under investigation.

Information on the frequency of exposure and content preferences was obtained on six media: radio, television, movies, comic books, books, and comic strips. Subsequent analysis showed that the pictorial media—television, movies, and comic books—form a unit in opposition to the others.[3] This fact, and the greater concern with the problem of these media than with some of the others, led to the decision to limit most of the analysis in this study to these three.

Three main indices of the social characteristics of the children's families are used: occupation of father, education of parents, and religion.

The main psychological emphasis in this study is on the consciously perceived problems a child has about himself and his relations with the people in his immediate environment. Does a child with such problems expose himself more to the mass media than children without them? Once exposed, does he use the media for different purposes than those who do not have these problems? Is he, because of his problems, affected in a different way by the media? Data were also collected on the child's tendency to be extrapunitive or intropunitive and on evidences of rebellious independence.

A subsample of the boys who filled in the questionnaire was interviewed individually, at which time they were given a comic book to read and asked in detail about their reactions to its content, what they remembered of it, how they perceived the people and events in it. The comic book used contains elements related to the particular thought processes involved in the study. It consists of three stories. The first concerns a young boy who was brought up under adverse circumstances and turns to a life of crime. He kills six people, robs a bank, and is finally caught and executed. The second story is about a scientist who has invented a way of getting into space in a very short time. The spacemen, on learning of this discovery, send an agent to earth to kill the scientist. Unknown to the scientist, his dog, whom he was about to return to the kennels, attacks the spaceman and saves the scientist's life. The third story features an individual named Strongman, who, by building up his muscles, has achieved enormous physical strength. He is called in by the police to investigate a dock murder only to discover that it is part of a Communist plot. Strongman finally saves the city from an atomic bomb by boarding a ship and vanquishing the Communists, a task the police and FBI had been afraid to perform.

With this same subsample, and at the same session, an experiment was conducted using the questions on the cognitive variables in the questionnaire as a premeasure, the comic book as an experimental stimulus, and repetition of the cognitive questions after the reading of the comic book as a postmeasure. This experiment furnishes the data for the imputation of effects. If one has isolated certain cognitive correlates of high exposure among certain types of children, one may investigate, by means of such an experiment, the effect of a particular example of the mass media on these cognitive factors among children with similar psychological characteristics who are not habitually very much exposed. The expectation is that even such a small degree of exposure may bring this group closer, with regard to the cognitive factor in question, to the psychologically equivalent group with high exposure, if this way of thinking is indeed an effect of the mass media.

PSYCHOLOGICAL AND COGNITIVE CHARACTERISTICS

What have here been called "psychological" characteristics are assumed to be more stable elements of the personality (more important for a child's general adjustment to his environment) than the "cognitive" factors. Within the present framework, this difference implies that the former would not logically be thought of as effects of exposure to the mass media; the latter might be. The psychological characteristics may more profitably be considered intervening factors, by means of which the relation between exposure and its determinants and effects may be elucidated. For this reason, the present section deals separately with the psychological and cognitive characteristics of the children in the sample. Its purpose is to describe the sample with regard to these characteristics, as well as to present relationships among them, and to outline their dependence, if any, on certain social factors.

---

[2] The questions dealing with psychological and cognitive characteristics are presented in Appendix B.

[3] Radio and books form another unit. Comic strip reading is mainly related to exposure to comic books. Confining analysis to media that correlate positively with each other insures greater homogeneity in different exposure groups. For this reason, comic strips—though also pictorial—were not among the media emphasized in this study.

Psychological Characteristics

Problems. We are concerned with the problems a child has in three areas: self, peer, and family.

Problems in the self area center on a child's expressed discontent about his abilities and appearance. They are indicated, for instance, by a child's belief that he is too fat, that he would like to be better looking, or smarter, or stronger. Problems in the peer area concern a child's feeling of inadequacy about his relations with other people of his age—for example, a feeling that he would like to have more friends than he has. A child who feels that his parents keep him from doing the things he wants to do or who feels that they do not understand him has problems in the family area. The total measure used for problems is the sum of the scores received in each of these areas, divided, for most of the analysis, into three categories: many problems, medium, few problems.

Each area has a minimum score of 0 which means no problems. The maximum score in the self area is 23; in the peer area, 26; in the family area, 27. The maximum total problem score is not the sum of the maxima for the three areas separately, since these maxima contain response patterns that are mutually exclusive. The total maximum is 54; the minimum is still 0.[4] Total scores received by the children do not reach either of these limits: they go from 1 to 43.

Table 1 gives the distribution of total problems for boys and for girls. Though boys and girls did not differ in reporting the presence of problems in general, the distribution in the three areas was not the same. The major difference was found in the family area, where boys have more problems than girls ($\chi^2 = 6.9$, $.005 < P < .01$).

TABLE 1

FREQUENCY AND PERCENTAGE IN EACH PROBLEM CATEGORY BY SEX

Problems	Boys		Girls		Total	
	Freq.	%	Freq.	%	Freq.	%
Many (≥ 21)	119	36	101	34	220	35
Medium (15–20)	87	26	75	26	162	26
Few (≤ 14)	126	38	118	40	244	39
Total	332	100	294	100	626	100

The presence of problems among boys is related to the occupation of the father and parental education. Boys whose fathers are workers are found to have higher problem scores than those of white-collar fathers ($\chi^2 = 5.7$, $.01 < P < .02$); yet sons of college educated parents have higher problem scores than sons of noncollege parents ($\chi^2 = 5.8$, $.01 < P < .02$). Further analysis shows that problems in the family area determine the latter relationship ($\chi^2 = 5.8$, $.01 < P < .02$).

Among the girls there is no significant relationship between problem scores and any of these social factors taken independently. However, among girls whose fathers are workers, Catholics have more problems than Protestants ($\chi^2 = 4.1$, $.025 < P < .05$). The role of religion is seen especially clearly in the self area. Catholic girls, no matter what the occupational level of their fathers, have more self-problems than do Protestant girls ($\chi^2 = 3.9$, $.025 < P < .05$).

Direction of punitiveness. Intropunitive children are those who tend to blame themselves in situations of conflict or frustration; extrapunitive children tend to place the blame elsewhere.

Scores on direction of punitiveness were based on five items in the questionnaire. Extrapunitive answers were given high scores; intropunitive ones, low scores.[5] The scores range theoretically from 1 to 19; actual scores received reach these limits.

Table 2 presents the basic data on direction of punitiveness. As can be seen from the extreme categories of this table, girls are very much more intropunitive than boys ($\chi^2 = 13.3$, $P < .001$). Direction of punitiveness is independent of all social factors under discussion here, for both boys and girls.

Rebellious independence. Ratings of rebellious independence were based on answers to the following question:

---

[4] The exact scoring scheme is given in Appendix C.

[5] The exact scoring scheme is given in Appendix C.

TABLE 2

PERCENTAGE IN EACH PUNITIVE CATEGORY BY SEX

Category	Boys (N = 332)	Girls (N = 294)	Total (N = 626)
Extrapunitive (≥ 8)	32	20	26
Medium (5–7)	41	41	41
Intropunitive (≤ 4)	27	39	33

If you were going to the circus would you want to go—

A—with your father?
B—with your best friend?
C—with your brothers or sisters?
D—with a group of friends?
E—with your mother?
F—by yourself?

Each child was asked to give his first, second, and third choice.

The children's answers fell into distinct groups: those children marking mother, father, siblings, as opposed to those marking, for instance, group of friends, best friend, and alone. The rejection of parents as circus accompaniers or the desire to go "by yourself" when expressed by 10- or 11-year-old children was considered indicative of an attitude of rebellious independence.

The scoring scheme follows:

Rebelliously independent: all of those children including "by yourself" as a choice, as well as those without this response who, however, do not choose either parent.

Nonrebellious: those children choosing mother, father, and siblings, as well as those children who do not have siblings hence do not choose this category, but who choose at least one parent and do not choose the "by yourself" category.

Medium: all other responses.

Table 3 presents the data on this question. Again there is a striking difference between girls and boys: many more boys than girls are rebelliously independent ($\chi^2 = 14.2$, $P < .001$). There is no relationship between rebellious independence of the child and the social characteristics of his family.

TABLE 3

PERCENTAGE IN EACH REBELLIOUS INDEPENDENCE CATEGORY BY SEX

Category	Boys (N = 332)	Girls (N = 294)	Total (N = 626)
Rebelliously independent	19	9	14
Medium	42	41	42
Nonrebellious	39	49	43
No answer	0 (1 boy)	1 (3 girls)	1 (4 children)

To summarize: girls and boys have about the same problem level. Boys, however, have more problems in the family area. On the other two psychological variables a more decided sex difference exists: girls are more intropunitive and less rebelliously independent.

When the interrelation among these three psychological factors is considered, only one significant relationship emerges: the rebelliously independent children are more extrapunitive, the nonrebellious ones more intropunitive ($\chi^2 = 12.2$, $P < .001$), a relationship particularly strong among the boys.

Cognitive Attributes

Stereotyping. Degree of stereotyping in the classification of people was gauged from four different stereotypes: good–bad, strong–weak, lazy–hard working, and honest–dishonest. The average number of stereotyped responses given out of 12 is 3.72, 3.86 for boys and 3.55 for girls.

This score is based on Questions 1, 2, 4, 6, 7, 9, 11, 12, 14, 16, 17, 19 of Part II of the questionnaire. The good–bad and strong–weak stereotypes were included in four items each, the others in two. More items were included for these classification principles because it was felt that they are more closely related to the content of the mass media. A score of 1 was given to the stereotyped answer, 0 to the nonstereotyped one. Hence, scores range theoretically from 0 (unstereotyped) to 12 (stereotyped). The actual range obtained was from 0 to 11. Children who left only one or two of these items unanswered were given a score by assigning an answer to the missing questions by means of the responses that were given and the over-all marginals on the questions. Six children (five boys and one girl), however, left so many of these questions unanswered that no over-all score could be given to them with any degree of certainty. Hence, the discussion of stereotyping is based on 620 children. For some of the analysis a three-category classification of stereotyping scores was used:

High: ≥ 6
Medium: 3–5
Low: ≤ 2

The four items to which the most stereotyped answers were given are:

(1) Do you think it is important to know if a person is good or bad before having anything to do with him?

I think it is very important—86%

(2) Lazy people are all alike and quite different from people who work hard.

I think this is true—72%

(3) Why do some people become bums?

Because they are naturally lazy—48%

(4) It is possible to tell by a person's face whether he is honest or dishonest.

I think this is true—37%

The strong–weak classification is thus not included among the questions drawing the most stereotyped responses. Both items of the lazy–hard working stereotype appear among the top four.

Amount of stereotyping is related to religion for both girls and boys. Catholics turn out to be more stereotyped than Protestants (boys: $\chi^2 = 8.0$, $.001 < P < .005$; girls: $\chi^2 = 3.9$, $.025 < P < .05$). The occupation of the father affects stereotyping only in the case of the boys: sons of worker fathers are more stereotyped than are those whose fathers have white-collar occupations ($\chi^2 = 5.0$, $P = .025$).

Perception of threat. The mean score of threat perception is 18.83 out of a possible 32.

This score is based on Questions 3, 5, 8, 10, 13, 15, 18, 20 of Part II of the questionnaire. A 1 was given to the response indicating least perception of threat, a 4 to the one indicating most perception; 2 and 3 were given to the intermediate response categories. Scores, therefore, can range from 4—least perception of threat—to 32—most perception. Actually, the range was from 9 to 27. Three children (all boys) were not included because of insufficient answers; the sample is, therefore, reduced to 623—329 boys and 294 girls—for this analysis. A three-category classification was used for some of the analysis:

Perception of threat—total: H ≥ 21
M 18–20
L ≤ 17
Personal: H ≥ 10
M 8–9
L ≤ 7
World: H ≥ 12
M 10–11
L ≤ 9

The average score for boys was found to be 19.23, for girls 18.38: boys perceive the environment as more threatening than girls do ($t = 3.6$, $P < .001$). The item that drew the most threat responses is:

Do you think something may destroy the world some day?

Yes, I definitely think so—19%

I think it might happen—47%

Perception of threat was divided into two categories: first, the perception of threat to the individual, a feeling of personal frustration; and second, perception of threat to the country or world. The item mentioned above as drawing the most threat responses is part of the second, or world, area. The item in the personal category that drew the most threat responses is:

Is it hard for a person to be what he wants to be in life?

It is very hard—15%

It is quite hard—42%

In the personal area the difference between girls and boys is not significant. Hence, the fact that boys perceive more threat in general is based on the fact that in the world category they give more threat responses than girls do ($t = 4.1$, $P < .001$).

Religion is related to amount of perceived threat to the world or country (world area) in the case of boys. Catholic boys see more of this threat ($\chi^2 = 4.2$, $.025 < P < .05$).

The frequency of problems among girls is related to the personal category. Girls with many problems tend to feel more personal frustrations than girls with few ($\chi^2 = 4.6$, $.025 < P < .05$). Among the boys, the rebelliously independent ones see more threat to the world and to themselves as individuals than the nonrebellious ones ($\chi^2 = 4.9$, $.025 < P < .05$).

Projected self-image, realistic. All children were asked what they would like to be, occupationally speaking, when they are grown up.

Each child was asked to give three occupational choices. The analysis, however, is based only on the first choice. The realism score was based on all three choices; hence, these percentages are based on only about two thirds of the total cases, those where realism of the first choice is clear.[6]

Tables 4 and 5 give the three categories into which most of the choices fall. The modal category for the boys, top professional, includes professions like doctor, lawyer, architect, professor. It does not in-

---

[6] Appendix C gives the exact coding scheme.

TABLE 4

Most Frequently Chosen Occupational Categories of Boys

Category	No. 1st Choices	%[a]
Top professional	52	16
Athletics	38	12
Worker (all skills)	37	12
Total (top 3 categories)	127	40

[a] Percentages are based on 320 cases; 12 boys did not answer this question.

TABLE 5

Most Frequently Chosen Occupational Categories of Girls

Category	No. 1st Choices	%[a]
Lesser professional	70	24
Teacher	52	18
Clerical and sales	49	17
Total (top 3 categories)	171	59

[a] Percentages are based on 293 cases; 1 girl did not answer this question.

clude engineer, which was classified separately, or scientist, which also had a separate category.

The top category for the girls, lesser professional, includes as the main occupation nurse. Teacher was separated from the other lesser professions because it appeared so frequently. The third most frequently chosen category of the girls, clerical and sales, has secretary as its main occupation.

It may be seen that girls show a more uniform pattern than boys. Over half of the girls' first occupational choices fall into the three most frequently chosen categories; just a little over a third of the boys' choices do. The girls of this age do not seem to let their imaginations roam as much in choosing their desired occupation. This uniformity is further brought out when the reality of these occupational desires is investigated. Only 37% of the boys' first choices, where realism was ascertainable, are realistic as opposed to 67% of the girls' choices.

How is the nature of the first occupational choice related to the social and psychological factors? Among boys it was found that first occupational choice is very much determined by their parents' position in society. The worker category, which is the third over-all category chosen, does not appear among the top three choices of Protestants, of boys whose parents went to college, or of boys with white-collar fathers. Among sons of worker fathers and sons of noncollege parents it is the modal choice.

Rebelliously independent boys introduce two new categories as their first and second most frequently chosen ones. The first refers to noncombatant air, sea, and other adventure occupations. The second is the armed services. Nonrebellious boys, in contrast, retain the modal top professional category.

The social characteristics of the girls do not affect their occupational choices. Lesser professional, teacher, and clerical and sales appear as the three most popular choices in all social subgroups. However, extrapunitive girls introduce a new category, which refers to occupations like actress and singer, and takes the place of clerical and sales in this subgroup.

Projected self-image, fantasy. Information on the projected self-image of the children in a fantasy setting was based on a question which presented them with a list of characters from which to choose the ones they would most want to be changed into (Question 26 of Part 1 of the questionnaire). Tables 6 and 7 present the three personalities most frequently chosen by the children in their first choices.

The greater uniformity shown by the girls than by the boys is again apparent. Almost three fourths of the girls' first choices are in the three most frequently chosen categories, as opposed to only a little more than a third of the boys'.

TABLE 6

Favorite Fantasy Metamorphoses of Boys

Person Changed Into	No. 1st Choices	%[a]
Superman	46	14
Animal breeder	39	12
FBI agent	37	11
Total (top 3 categories)	122	37

[a] Percentages are based on 331 cases; 1 boy did not answer this question.

TABLE 7

Favorite Fantasy Metamorphoses of Girls

Person Changed Into	No. 1st Choices	%[a]
Movie star	132	45
Princess	44	15
Animal breeder	37	13
Total (top 3 categories)	213	73

[a] Percentages are based on 291 cases; 3 girls did not answer this question.

Animal breeder, for boys, seems to be a high status choice. It occurs first or second among Protestants, boys whose parents went to college, and boys with white-collar fathers. It does not come up in the first three choices of Catholics, boys whose parents did not go to college, or sons of worker fathers. These latter three groups substitute movie star for animal breeder; movie star is in first or second place for all of them. The only other change of interest is that the group of boys whose parents went to college has president as its modal choice, a selection that does not occur in even the top three categories of any other social subgroup. This change is also made by boys who have few problems in the peer area; they also choose president, a different type of power personality, more frequently than Superman.

The girls are so definitely decided on movie star as their main desire that there is no variation among subgroups—each subgroup has this selection as its modal choice. The second and third favorite categories of the girls, princess and animal breeder, are represented among the top three choices of all subgroups except the Catholics, among whom reporter takes the place of animal breeder, and among girls with many peer problems where cowboy, reporter, or queen takes the place of animal breeder.

Passivity, professed activities. One approach to the problem of passiveness centered on the preference for more or less active forms of social behavior.

This score was based on Questions 1, 4, 18, 22, 24 of Part I of the questionnaire. An active choice was given a score of 1, a passive choice 0. Hence, scores can, and do, range from 0 to 5. Two hundred and seventy boys and 266 girls answered all five questions. The rest are not included in the discussion.

Given five choice situations, girls on the average chose 2.94 active alternatives as opposed to 2.65 for the boys, a significant difference ($t = 3.4$, $P < .001$). Rebelliously independent girls gave fewer active choices than nonrebellious ones ($t = 2.3$, $.02 < P < .05$).

Passivity, change in occupational status. An alternative approach was through occupational aspirations. Some of the occupational choices of the children were realistic, others were not. Of the former, certain ones were considered active—surpassing the occupational status of the parents.[7] Of boys' realistic occupational choices, 30% indicate an aspiration for an occupation higher on the social scale than that of the father. Girls' choices, which were compared with the occupation of the mother, are much more active: 74% of their realistic choices are active in this sense, mainly because a large proportion of the mothers were housewives, which made any occupational choice an active one. These tendencies are independent of the social and psychological characteristics under investigation.

Exposure Habits

The present section concerns the frequency of exposure and content preferences of the children in the sample, and the relation of these exposure habits to social and psychological characteristics. In interpreting the data of this section, it must be remembered that information on frequency of exposure was obtained from the reports of children. No validation of this information was attempted. However, the average exposure times reported agree quite well with results of other studies (e.g., Butterworth & Thompson, 1951; Mahony, 1953; Mit-

---

[7] In some cases a realistic choice could not be classified into either of these categories; such a choice was considered unascertainable in this respect. Only the number of realistic choices that it was possible to classify into the active or nonactive category was used as a base. Appendix C gives the exact coding scheme.

TABLE 8

MEANS OF FREQUENCY OF EXPOSURE FOR BOYS AND GIRLS

Media	Boys		Girls	
	Mean	Classification	Mean	Classification
Television[a]	4.7	Between "a couple of hours a day" and "3 or 4 hours a day"	4.6	Same as for boys
Comic books	3.4	Between "1 every few days" and "about 1 a day"	2.7	Between "1 every few weeks" and "1 every few days"
Comic strips	4.5	Between "read them a few times a week" and "read a few every day"	4.3	Same as for boys
Radio	2.7	Between "a little every week or so" and "a little every few days"	2.6	Same as for boys
Movies	2.6	Between "about once a month" and "a couple of times a month"	2.2	Same as for boys
Books	3.8	Between "1 every few months" and "a couple a month"	4.4	Between "a couple a month" and "1 a week or so"

Note.—Means are based on a 6-point frequency distribution.

[a] Twenty-five children, 13 boys and 12 girls, do not have television sets in their homes. Of these, about half (56%) watch television rarely, the rest are nonetheless exposed, probably at friends' houses.

chell, 1929; Witty, 1952). Moreover, certain investigators—who have checked validity—have shown that children's reports are fairly correct (e.g., Dale, 1935).

Extent of Exposure

Table 8 presents the mean exposure to each of the six media investigated in this study. It shows that television takes up by far the most exposure time of children; books and movies least. The means for girls and boys fall into the same frequency categories for television, comic strips, radio, and movies. For books and comic books, however, boys' and girls' averages fall into different, though adjacent, categories. Girls read fewer comic books and more books than do boys.

Do children with high exposure to one medium expose themselves to other media, or is their exposure time reserved for the favorite medium? The correlations among frequencies of exposure to the media, which will answer this question, are given in Table 9. The upper half (to the right of the diagonal) represents the correlations for boys, the lower half those for girls.

Two clusters of media emerge when these correlations are examined: one consists of the pictorial media—television, movies, and comic books—frequencies of exposure to which correlate positively with each other; the other consists of radio and books. The amount of exposure to neither of the latter group—radio and books—correlates positively with any of the former; the frequencies of both correlate negatively with amount of television viewing.

As has been indicated in the introductory section, this study centers on the first cluster of media—television, movies, and books.

The exposure score for these media was derived as indicated in Appendix C. The minimum score is 16, which represents a child who "hardly ever" watches television, "hardly ever" goes to the movies, and "hardly ever" reads comic books. The maximum score is 72 which would be given to the child who watches television "5 to 6 hours a day or more," goes to the movies "almost every day," and reads "4 or 5" comic books "a day or more." The scores actually reach both theoretical limits. For most of the analysis a three-category classification of exposure was used:

High: ≥50
Medium: 42–49
Low: ≤41

The difference between the mean of the high exposure group and that of the low exposure group, for both boys and girls, is 2.1 standard deviations of the total distribution. The high exposure boys have a mean of 57.34, the girls a mean of 56.15. The low exposure groups have the means of 33.11 and 32.68, respectively, for boys and girls. Taken as two separate samples—one high in exposure, one

TABLE 9

Correlations Among Media for Boys and Girls

Media	Television	Movies	Comic Books	Radio	Books	Comic Strips
Television		.29*	.41*	−.16*	−.17*	.09
Movies	.24*		.34*	.04	−.02	.09
Comic books	.34*	.30*		.02	−.01	.21*
Radio	−.14*	.02	.06		.15*	.05
Books	−.46*	−.03	.05	.16*		−.02
Comic strips	.06	.04	.21*	.18*	.07	

Note.—Upper half represents correlations for boys; lower half, correlations for girls.
* $P \leq .05$.

low—these differences are highly significant. Hence, one can say that the high and low exposure groups really are different in exposure. Their means differ by about 24 points, for both boys and girls, on a distribution with a range of 56.

The mean exposure score for boys is 45.23, for girls it is 41.20, a significant difference ($t = 4.4$, $P < .001$). As can be seen from Table 8, this difference is largely, though not completely, dependent on the difference between boys and girls in the frequency with which they read comic books.

High exposure occurs more frequently among Catholic children than among Protestants ($\chi^2 = 4.8$, $.025 < P < .05$) and among those whose fathers are workers rather than those whose fathers have white-collar occupations ($\chi^2 = 8.0$, $.001 < P < .005$). However, the degree of association that exists between these social characteristics and the child's frequency of exposure is not very large. Neither occupation of father nor religion accounts for very much of the variation in degree of exposure: father's occupation accounts for 1.7% ($r_m = .13$), religion for 1.0% ($r_m = .10$).[8]

More important than these factors are two others: parental restriction on amount of exposure and IQ.[9] The former, which accounts for 15.9% ($r_m = .40$) of the variation in degree of exposure to the pictorial media, may be considered an over-all manifestation of the family's social and cultural standing. The latter works independently of these other factors and accounts for 7.7% ($r_m = .28$) of the exposure variation.

It may be of interest to note, at this point, that though grades and reading ability seem to relate to amount of exposure, IQ differences explain both of these relationships.[10]

There are, then, four social factors that affect the amount of exposure of children.

---

[8] The statistic used—r_m—is computed from the frequencies of 3 × 2 tables: 3 levels of exposure; 2 of the social characteristic. The subscript "m" is used to distinguish this measure of relationship from the usual correlation coefficient; it stands for Mosteller, who devised it (see Lazarsfeld & Rosenberg, 1955, p. 74). Though degree of relationship is measured by means of r_m in this section, significance of relations is determined by means of χ^2 based on the extreme exposure groups, i.e., on a comparison between the high and low exposure groups with respect to the particular factor under investigation.

[9] The IQ scores are based on the California Elementary Mental Maturity Test—Short Form, 1950. "High" includes scores of 123 or more, "low" those 108 or less, and "medium" those falling between 109 and 122. Analysis is based on the extreme groups.

Three levels of restriction are used. The restricted group consists of children for whom none of the pictorial media—movies, television, comic books—is unrestricted. The unrestricted group consists of those children who are restricted neither in television nor in comic books. Since movies are so frequently restricted, such a restriction was allowed in this group if accompanied by a lack of restriction on radio. In other words, this group consists of children whose parents do not restrict movies, television, and comic books, as well as those who do restrict movies but do not restrict television, comic books, and radio. All other children fall into a middle category. The two extreme groups are used in the analysis.

[10] Grades in substantive subjects were averaged after assigning scores of 1 to A, 2 to B, etc. "High" consists of those averages that were 1.85 or less, or a little over B or above. "Low" were those that were 2.65 or more, that is, C+ or below. "Medium" falls between these two points.

The relationship between reading and exposure disappears completely when measured separately

TABLE 10

COMBINATION OF THE FOUR DETERMINING FACTORS (AMOUNT OF RESTRICTION, IQ, OCCUPATION OF FATHER, RELIGION) AND EXPOSURE FOR TOTAL SAMPLE

Factor Combinations	Exposure			Total
	High	Medium	Low	
Restricted, High IQ, White Collar, Protestant[a]	5%	19%	76%	100% ($N = 21$)
Unrestricted, Low IQ, Worker, Catholic	63%	23%	14%[b]	100% ($N = 22$)

[a] Actually 28 children have all these characteristics. However, 7 of these—all with low exposure—do not have television sets at home and hence are excluded from the table.
[b] All the children in this group are girls.

Two of them—religion and occupation of father—are background characteristics; they relate least to amount of exposure. Amount of restriction, which accounts for most of the variation in exposure, represents the direct influence of parents. The child's IQ seems to work as a selective factor: the high IQ child is still less exposed than the low IQ child even when both relevant social characteristics as well as parents' influence do not work against exposure, as in the case of the unrestricted, worker, Catholic group ($r_m = .22$).

These four factors affect exposure independently of each other. Any combination of them increases one's ability to account for variations in exposure. A combination of all four factors accounts for 47.1% of this variation. This information is presented in Table 10, even though such a procedure naturally reduces the number of cases greatly.

A new element is introduced when the relationship between achievement and IQ is considered for *each individual child.* Each child was rated as an over-achiever (school grades exceed IQ), an under-achiever (grades fall below IQ), or as a normal achiever.[11] Normal achievers fall into exposure classifications with expected frequencies—expected in the sense that they are similar to those for the sample as a whole. Over- and under-achievers deviate from this pattern, and in this deviation there is a decided difference between boys and girls.

There is no relation between achievement relative to IQ and exposure among the boys. Among the girls, however, a relationship does exist. Girls who are over-achievers are less exposed than girls who are under-achievers ($\chi^2 = 6.2$, $.01 < P < .02$). It seems, then, that girls who take school work seriously, who are over-achievers, spend less time in exposure to the mass media than do the girls who are under-achievers.

Exposure and psychological characteristics. An obvious question to ask is whether having problems leads a child to expose himself more to the mass media. Will the mass media be sought more by children with many problems in their everyday life, perhaps as an escape from these problems? The answer seems to be No: presence of problems is not related to amount of exposure to the pictorial media. Problems do, however, correlate with exposure to the nonpictorial media. Children with many problems tend to listen *less* to the radio and read *fewer* books than do chil-

---

among children with high and those with low IQ. The same occurs with respect to achievement in the high IQ group. In the low IQ group a tendency remains for children with low averages to be more exposed than those with high averages. However, the χ^2 value of this relationship is not statistically significant and the degree of the relation is reduced from .31 for the sample as a whole to .18 for this group.

[11] Both IQ and average grades were classified into ten categories. If the two classifications for a given person differed by at most one point, that person was rated as a normal achiever; if IQ fell into a category two or more points higher than average grades a rating of under-achievement was given; if IQ was two or more points lower than grade average the child was considered an over-achiever. Girls' and boys' IQ scores have about the same distribution. Girls, however, have much higher grades; hence, there are more over-achievers among them (25%) than there are among boys (9%).

All children in the top or bottom two IQ categories were eliminated in the analysis because of the operation of a ceiling effect. Children whose IQ's are in the top two categories cannot be rated over-achievers, because the category of their grades cannot exceed that of their IQ. Similarly, children with IQ's in the lowest two categories cannot be rated under-achievers.

dren with few problems (boys: $r = .22$; girls: $r = .14$).[12]

Among the boys, exposure is unrelated to direction of punitiveness. Among the girls, however, the ones who are extrapunitive tend to be somewhat more exposed ($\chi^2 = 3.8$, $P = .05$). There is no relation between exposure and rebellious independence for either boys or girls.

Though problems and exposure to the pictorial media are not related directly, the presence of problems may affect the relationship between exposure and other variables. For example, high IQ children are less exposed to the mass media. It might be that such a relationship exists only among children with many problems, whereas in the group with few problems, IQ and exposure are independent.

Of the four factors most related to exposure there is a conditional relation only in the case of father's occupation. Among children with many problems, occupation of father is barely related to exposure ($\chi^2 = 1.3$, $P = .25$, $r_m = .10$). In contrast, among children with few problems, those whose fathers are workers are much more exposed than those whose fathers hold white-collar positions ($\chi^2 = 7.5$, $.005 < P < .01$, $r_m = .21$). In other words, among children with few problems father's occupation is more determinant of exposure than it is in the group with many problems. Perhaps a child with many problems is less likely to accept family standards—less emphasis on the mass media in white-collar families, for instance—than one with few problems.

There is, however, another way in which problems may be related to exposure. Problems and some other characteristic may jointly affect the amount of exposure, though neither factor alone may do so. An example of this type of relationship concerns the degree of rebellious independence of the child. Among children with few problems, independence and exposure are not related at all. A fairly strong relationship exists, though, in the group with many problems. In this group, children who show rebelliously independent tendencies are more likely to fall into the high exposure group than are those who do not show such rebelliousness ($\chi^2 = 9.5$, $.001 < P < .005$, $r_m = .31$). Since degree of rebellious independence has no direct relationship to amount of exposure, this means that children with many problems who *also* show a certain amount of rebellious independence tend to be more exposed to the pictorial media, though neither factor alone produces significant variations in exposure.

Content Preferences[13]

In order to investigate content preferences, it is necessary to devise a scheme for classifying the content of the mass media. Some categories are applicable to all media, others only to some. Westerns, crime stories, animal tales, comedies, plots centering on spies and war, and horror stories are descriptive of material on any medium. So are stories centering on everyday family themes and the category "superforce," referring to heroes—like Superman—endowed with supernatural powers. Other classifications apply particularly to certain of the media and not to others: as for instance sports, music, and religion on radio and television, and cartoons in the movies. It is in such categories that the stated preferences of the children were classified.

A percentage score was arrived at by the following procedure: only the first three choices of each child for television, movies, and comic books were used, since many children did not fill in all five choices. A first choice was given the score of 3; a second that of 2; a third, 1. Hence, there is a maximum score of 6 for each medium. The three pictorial media were treated as a unit in this analysis, giving a maximum possible score of 18 for a child who filled in at least three choices for each medium. The crime choices, for instance, were scored in the same way. The child whose first choice on television is *Dragnet* and whose third choice in comic books is *Crime Does Not Pay,* and who mentions no other crime material, would get a score of 4 on crime (3 for first choice on television plus 1 for third choice on comics). If the child had indeed given three choices for each medium, his percentage score for crime would be 4/18 = 22%. Had he, however, given no movie preferences, his total maximum score would be reduced to 12, and for the same pattern of crime choices he would get a score of 4/12 = 33%. In this manner each child was given a percentage score on each of the eight categories that covers all three media; a score that represents the percentage the scores in a particular category are of his maximum possible score. It is the means of these scores that are used for analysis. Seven children, three boys and four girls, gave no content preferences for any of the pictorial media. Hence, the analysis is based on 329 boys and 290 girls.

---

[12] Scores for exposure to radio and books were formed by adding up the number of the frequency given for each medium. These numbers went from 1 to 6, 1 representing a high frequency of exposure, 6 a low one. Therefore, the combined scores ranged from 2 to 12. It was these scores that were correlated with problem scores. Only children who answered both the radio and book frequency questions were used in this correlation: 323 boys, 291 girls. Since high scores on frequency represent low amount of exposure, the sign of these correlations has to be reversed.

[13] This information is based on the children's write-in answers to questions of the form, "Write here the names of the 5 programs [movies, comic books] you like best" from Part III of the questionnaire. Each child was urged to fill in only as many as he really liked, even if this number was fewer than five.

The rubrics applicable to the pictorial media are listed with the frequencies of their choice in Table 11. It can be seen from this table that animal and situational stories comprise a large proportion of children's favorite material, especially for girls; almost half of their choices fall into the latter classification.

Five of the categories presented in Table 11—western, crime, spy and war, superforce, space—seem to contain similar elements: they tend to involve heroes who defend themselves and their friends, their rights, or their country and the world against evil elements that threaten them by force; their defense usually involves acts of force as well. This cluster may be referred to as the aggressive-hero type of material, or AH content for short. This type of material is central to the present study for two reasons: First, it has greater social relevance—the concern about the effect of the mass media on children centers mainly on stories having these features; second, stories of this nature contain the elements specific to the thought processes under investigation.

TABLE 11

MEAN PERCENTAGE SCORES FOR CONTENT CLASSIFICATIONS BY SEX

Classification	Boys	Girls
Animal	22.06	19.08
Situational[a]	15.30	44.54
Western[b]	14.21	7.27
Crime	7.87	3.91
Spy and war	5.07	1.31
Superforce	4.37	2.36
Horror	3.90	1.09
Space	2.99	1.49
Other[c]	24.23	18.95
Total	100.00	100.00

[a] Refers to comedy, variety, and quiz programs; musicals and comedies; and comic book stories centering on family themes.

[b] Jungle stories are included in this category.

[c] Includes all categories that do not apply to all three media and choices that were impossible to classify.

Each child was given an AH score by taking the sum of his percentage scores for the five areas involved. The mean of these scores for boys is 34.51, for girls it is 16.34. This difference of 18.17 points is significant ($D = 18.17$, $t = 10.9$, $P < .001$).[14] Because of this difference—and others, as will be shown—boys and girls are discussed separately, starting with boys.

Boys' preference for AH content. What factors seem to determine content preferences of boys? Of the social characteristics—religion, father's occupation, parents' education—none affects liking for AH material significantly.

Restriction on exposure time, however—indicative of the child-rearing practices of the parents—is important for the type of material preferred, as it is for frequency of exposure. Boys who are not restricted on amount of exposure show greater preference for AH content ($D = 9.46$, $t = 3.1$, $.001 < P < .01$). There are two other indications of this part of the child's environment in this study: whether or not the child gets spanked and who in the family takes the role of the punisher. The first of these—presence of spanking—plays a role in content preferences, though the second does not. Boys who sometimes get spanked show greater liking for AH content than boys who do not get spanked ($D = 9.86$, $t = 3.5$, $P < .001$).

---

[14] For most of the analysis in this section, D—the difference of average percentage scores between two groups—is used to indicate the relative importance of the factor under consideration, similar to the use of r_m in the previous section. A t test is used when the statistical significance of such a difference is considered.

IQ, which, as has been shown, is important for amount of exposure, also relates to differences in content preferences. Boys with low IQ's show greater preference for AH material than do high IQ boys ($D = 8.18$, $t = 2.6$, $.001 < P < .01$). With respect to achievement relative to IQ, it turns out that boys who are over-achievers—whose grades exceed their IQ—tend to show more liking for this type of material than boys who are under-achievers ($D = 10.05$, $t = 2.5$, $.01 < P < .02$).[15]

What of the psychological variables of this study? How do they relate to content preferences? Neither degree of problems nor direction of punitiveness affects preference for AH material among boys. Rebellious independence, on the other hand, shows such a relation. Boys who are rebelliously independent show more liking for AH content than do nonrebellious boys ($D = 10.13$, $t = 2.9$, $.001 < P < .01$). The stated desire of boys of this age to go alone to the circus seems to be accompanied by liking for this type of material.

There are then, four factors that affect liking of AH content among boys. On the basis of the amount of difference they make in preference for this type of material, they can be ranked in order of importance: rebellious independence, spanking, amount of restriction, and, least of all, IQ. The relation of achievement to IQ also seems to show such a relationship, but because of the small number of cases the conclusion here remains more tentative. It should be pointed out that these factors are related to AH preference independently of each other.

Even though problems show no direct relation to content preferences, it is possible that the relation between some of these factors and preference for AH content differs in the different problem groups. One of these elements—IQ—is conditional on the problem level of the child. Among boys with many problems, those with low IQ's show greater preference for AH content than those whose IQ's are high ($D = 8.46$). In the group with few problems, however, the picture changes. In this group IQ no longer makes much difference ($D = 2.36$): the low IQ child who has few problems does not show much more liking for AH material than the corresponding high IQ boy.

To summarize: boys who are rebelliously independent, get spanked, are not restricted on exposure time, and have low IQ's show particular preference for AH content. Further, IQ is mainly determining of the content preferences of children with many problems.

These factors indicate that content preferences of boys are somewhat more related to psychological characteristics than is the amount to which they are exposed.

Frequency of exposure and content preferences of boys. Boys who are much exposed to the pictorial media show greater preference for AH content than do those with low exposure ($D = 11.53$, $t = 3.8$, $P < .001$). Hence, more than amount of exposure differentiates the high and low groups; these groups also differ in their preferred content. This relationship can be interpreted in a number of ways. High exposure may influence a boy in such a way that he will come to prefer content of this type. Conversely, it is possible that it is the children with a particular liking for AH material who will expose themselves most. On the basis of the material in this study, one cannot prove that one of these possibilities is right, the other wrong. One can, however, investigate the relationship, and on the basis of such an analysis make an inference as to which seems more plausible. First, however, it is necessary to take into account another alternative: the relation might be explained by the joint effect of a third factor on amount of exposure as well as on preference for AH content.

Two factors—parental restriction of ex-

---

[15] Because of the ceiling effect mentioned in Footnote 11, only boys whose IQ's do not fall into the top or bottom two categories are used in this comparison. This limitation leaves only 16 over-achieving boys; hence, the result is tentative at best.

TABLE 12

Boys' Mean Percentage Scores of AH Content by Exposure and Rebellious Independence

Exposure	Rebelliously Independent	Non-rebellious
High	46.45 (N = 31)	32.45 (N = 40)
Low	31.65 (N = 20)	26.65 (N = 46)

posure time and the child's IQ—are associated both with exposure and with content preferences. Either might explain the relationship under consideration. Neither, however, does. Boys with high exposure show more preference for AH content when they have high IQ's as well as when their IQ's are low. Similarly, boys whose exposure time is not restricted show greater preference for material of the AH variety when they have high exposure. In the restricted group, on the other hand, this difference disappears. Perhaps children who are highly exposed to the pictorial media, despite parental restriction, do not increase this possible area of conflict further by preferring material disliked by their parents. Since, however, the relation between amount of exposure and preference for AH content remains in the unrestricted group, the factor of parental restriction cannot be said to explain this relationship. The analysis of the nature of this association, therefore, remains.

One way of approaching this question is to investigate the interaction of exposure and the psychological characteristics of the child with regard to his content preferences. One of these psychological factors—rebellious independence—is independently related to content preferences. Table 12 gives the joint relationship of this element and amount of exposure to liking for AH content. As is shown by this table, rebelliously independent boys who are highly exposed to the pictorial media are the ones who show greatest preference for AH content—more than would be expected by the mere addition of the influence of each of these two factors.

TABLE 13

Boys' Mean Percentage Scores of AH Content by Exposure and Problems

Exposure	Problems	
	Many	Few
High	44.30 (N = 40)	33.25 (N = 48)
Low	28.33 (N = 42)	27.64 (N = 47)

TABLE 14

Boys' Mean Percentage Scores of AH Content by Exposure and Direction of Punitiveness

Exposure	Extrapunitive	Intropunitive
High	43.82 (N = 38)	33.86 (N = 35)
Low	28.82 (N = 34)	33.18 (N = 34)

A similar phenomenon is apparent when AH preferences are looked at for children in different exposure-problem groups, as can be seen in Table 13. Again it is the combination of many problems and high exposure that is most associated with a great deal of liking for AH material.

Table 14 shows how the effect of exposure on content preferences relates to direction of punitiveness. Here, too, it is the boys with extrapunitive tendencies and high exposure who show particular preference for this type of material.

As a digression it is of interest to note that the direction of punitiveness of the child has a different effect on preference for AH content when the child has low exposure than when he has high exposure. If one accepts the premise that children who show great preference for AH material do in fact see more of it, one might speculate that extrapunitive boys use this type of material as an outlet for reactions that might otherwise result in somewhat antisocial behavior. Tentative confirmation of this interpretation comes from looking at the grades the boys have in conduct.[16] Among boys who get high grades in conduct, the extrapunitive ones have greater preference for AH material than do the intropunitive ones (D = 5.90). Among boys with low conduct grades, however, the relation is reversed: extrapunitive ones show less preference for AH content than intropunitive ones ($D = 12.10$).

---

[16] Boys who in their conduct grades for 1954–55 received at least one A were considered as having high conduct grades; those receiving nothing higher than a C were considered as the low conduct group. The results are not more than tentative because only 55 boys have low grades on conduct, a number which, when broken down by other variables, is not very large.

To summarize on the main point: boys who are highly exposed to the pictorial media are more attracted to AH content than are those with low exposure. This differential in content preferences is particularly apparent when high exposure is combined with problems in a child's life, extrapunitive leanings, or rebelliously independent tendencies.

These findings, combined with the result previously mentioned (that amount of exposure of boys is more socially determined, content preferences more psychologically), lend support to the following interpretation of the relation between exposure and content preferences: the combination of certain psychological characteristics in the child and high exposure to the pictorial media leads a boy to prefer the aggressive-hero type of material. This point will be considered more fully in the concluding section. Independent of interpretation, however, the above discussion shows that exposure groups are more complex than mere differences in frequency would indicate.

Girls' content preferences. It has already been mentioned that girls generally show little liking for this type of material. This fact is not particularly changed by any factor: no single element relates significantly to girls' preferences for AH content. Keeping this over-all uniformity in mind it might still be of interest to note the three factors that create the greatest differences in degree of liking for AH material among girls. They are, in order of importance, rebellious independence, direction of punitiveness, and IQ. Like the boys, rebelliously independent girls tend to show more liking for AH content than do nonrebellious ones ($D = 5.48$), and girls with low IQ scores show more than those with high IQ's ($D = 4.47$). Extrapunitive girls have a tendency to state more AH preferences than do intropunitive ones ($D = 4.91$).

Surprisingly, amount of exposure—which is not independently associated with AH preference—is quite strongly related to liking for AH content among girls with few problems. In this group, high exposure does seem to be accompanied by preference for material of the aggressive-hero type ($D = 14.96$). Further, extrapunitive tendencies and rebellious independence in girls are particularly strongly related to liking for AH material when accompanied by few problems (rebellious independence: $D = 11.71$; direction of punitiveness: $D = 20.55$).

The general conclusion must still be that girls' content preferences are not very much influenced by any of the factors under investigation. A number of relations, however, seem to be intensified in the group with few problems. As has already been indicated in the preceding section, girls generally show more uniformity than do boys; they seem to gravitate toward a norm. Perhaps it is only the girls with few problems who can afford to deviate from this norm and to be influenced fully by their own individual characteristics. Whether or not this interpretation is correct, one is left with the undeniable fact that as far as content preferences are concerned, girls show a much more uniform pattern than is shown by the boys.

REACTIONS TO A COMIC BOOK

The present section investigates the way children view a specific example of the mass media—a comic book. An attempt is made to isolate the behavioral and cognitive reactions to this comic book of children with high and low exposure and different psychological characteristics.

The data of the present section are based on interviews with 100 boys selected from the original sample on the basis of their degree of exposure and problem score, and equated, as closely as possible, on certain social characteristics and capabilities.[17]

---

[17] See Appendix A for a description of the selection of these groups. The distribution of the boys interviewed on these characteristics is available in the dissertation copy (Bailyn, 1956, p. 60). Genevieve Rogge Meyer shared with me the task of conducting these interviews; I am most grateful for her expert help.

Behavioral Aspects

Each boy was told to read the comic book in the same way he would if he were at home and to take as much or little time with it as he wanted. Unknown to him, the interviewer kept track of the time he spent on it. The average time spent was 19.1 minutes, with a range of from 5 to 47 minutes. The children with high exposure spent less time on the comic book—an average of 16.9 minutes—than did those with low exposure, whose average was 21.2 minutes ($D = 4.3$, $t = 2.3$, $.01 < P < .025$). This difference is intensified among children with high IQ's ($D = 6.0$). Among low IQ boys, whose reading is somewhat slower, exposure does not make quite as much of a difference ($D = 3.6$).[18]

The interviewer observed each child carefully—though not obviously—as he read the comic book. One point noticed was whether the child seemed to be reading the comic book or whether he mainly looked at the pictures. A little over half of the children (52%) seemed to be reading; somewhat over one fourth (27%) concentrated mainly on the pictures; of the rest (21%) it was either impossible to tell or both approaches were used by the child at different times. The children who read spent more time on the comic book than those who concentrated on the pictures. There was not much difference between high and low exposure children in this respect: children with high exposure showed somewhat more of a tendency to look mainly at the pictures than did those with low exposure. However, even when children who read the comic book are separated from those whose emphasis was on the pictures, high exposure boys spent less time on it:

	Read	*Pictures*
High exposure	21.9 min.	9.8 min.
Low exposure	26.1 min.	13.2 min.

TABLE 15

PERCENTAGE OF CHILDREN WITH DIFFERENT DEGREES OF INVOLVEMENT IN THE COMIC BOOK BY EXPOSURE

Exposure	Level of Involvement			Total
	High	Medium	Low	
High	52	36	12	100 ($N = 50$)
Low	26	50	24	100 ($N = 50$)

The interviewer also recorded the degree of the child's absorption in the comic book, his awareness of people and events around him, restlessness and other nervous gestures, expressions. On the basis of this information each child was assigned to one of three levels of involvement with the comic book.

High involvement: children who were rated as completely absorbed, showed no awareness of the experimenter or external events (with the exception of a few cases where an interruption occurred that had to be responded to), sat relatively still, and showed no overt signs of boredom, lack of interest, pure task orientation.

Low involvement: children who did not seem absorbed in the reading, showed awareness of external events, whose behavior indicated lack of interest, boredom.

Medium involvement: children whose behavior fell between these two groups.

The interviewer, of course, had no knowledge at the time of the interview of which exposure-problem group the child belonged in.

Over one third of the children (39%) were highly involved, 43% showed a medium degree of involvement, and 18% did not seem involved at all. Table 15 shows the degree of involvement of children with high and low exposure.

As can be seen from this table, children with high exposure showed much more involvement in the comic book than did those with little habitual exposure to the pictorial media ($\chi^2_{(2\times3)} = 7.5$, $.01 < P < .0125$). This difference between habitually high and low exposure children in the degree of their involvement in the comic book persists for children with many or few problems, those who are extrapunitive or intropunitive, and those with nonrebellious tendencies. Among the group of boys who are rebelliously independent, however, those with low exposure show a slightly higher degree of involvement than do those with high exposure.

---

[18] The high IQ children in the interview sample are divided fairly evenly into high and low exposure groups. The matching was less successful with regard to low IQ; these children are more heavily represented in the high exposure group, with the result that there are only 9 children with low exposure who have low IQ's.

There are, then, certain behavioral differences between children who are habitually highly exposed to the pictorial media and those who are not, when confronted with a comic book. High exposure boys spent less time on the comic book but became more involved in it than did boys with less exposure.

Cognitive Aspects

A comic book story is a unique thing; yet it has many points of similarity with other comic book stories. The question raised now is whether children with high exposure tend to view a particular story as an example of a type rather than as unique. Are they more attuned to the typical elements in the story, and less to the unique, than children with low exposure? As a result of acquaintance with this type of material on the one hand, and, on the other hand, of the use to which it is put, the expectation is that children with high exposure—perhaps only when accompanied by certain psychological characteristics—view the stories more as fitting into a general structure than as unique.

Each child who participated in the interview was asked to tell the interviewer the stories he had read. After each retelling the experimenter asked the child a number of specific questions dealing with the story.[19] If it is true that the child with high exposure views the stories more as a type than the low exposure child, he should remember fewer of these specific details. The average number of details remembered was 10.7 out of a possible 24, with a range of from 1 to 21. The mean for children with high exposure was 9.7, for those with low exposure, 11.6, a significant difference ($t = 2.3$, $.01 < P < .025$).

TABLE 16

AVERAGE NUMBER OF DETAILS REMEMBERED BY EXPOSURE AND INVOLVEMENT

Exposure	Involved	Uninvolved[a]
High	11.2 (N = 26)	8.2 (N = 24)
Low	12.0 (N = 13)	11.5 (N = 37)

[a] The medium and low levels of involvement have been combined to form this group.

There are, however, other factors that determine the number of details remembered by a child. The child's IQ and the amount of time spent on the comic book are perhaps the most relevant. One would expect a child with high IQ to remember more details than one with low IQ; and a child who spent more time on the comic to have a better recall of the elements in it than the child who read it very fast. Both of these expectations are confirmed. However, even when the amount of time is held constant, children with high exposure remember fewer details: adjusted means—high exposure, 10.0; low exposure, 11.4, $t = 1.7$, $.025 < P < .05$. It should be pointed out that the difference is especially evident among children who spend little time on the comic book; among those spending a great deal of time on it there is less difference between exposure groups. Neither does IQ account for this relationship. In the low IQ group, exposure does not affect the number of details remembered; among children with high IQ, however, the relation is much intensified.

Boys with low exposure do remember more of the details of a comic book story than do those with high exposure. These boys are also less involved in the comic book. Table 16 shows the number of details remembered by high and low exposure boys with varying degrees of involvement in the stories.

Two things are immediately apparent from the table. Involvement does not seem to affect the number of details remembered by children with low exposure, whereas it is an important factor for those with high exposure: boys in this group who are involved remember more details than do those uninvolved. Further, the table shows that children who are involved in the story remember about the same number of details from it whether they have high or low exposure.

On the basis of the hypothesis that children with high exposure are more attuned to the typical elements in the comic book than are those with low exposure, one would not expect the involved children with high exposure to remember as many details as those with low exposure, unless their memory centers around elements that occur in many comic book stories—around typical elements.

---

[19] Because of lack of space material from the interview session could not be included here. This material is available in the dissertation copy (Bailyn, 1956, App. B).

The third story the children saw—Strongman—is a good test case of this point.[20] Just as Strongman is about to subdue the Communists—the villains of this story—he is knocked out. Such an occurrence is not unique to this particular story. The hero often faces such a difficulty—it makes his eventual success more dramatic. If children with high exposure remember such a typical detail—frequently, as in this case, aggressive—more often than children with low exposure, and if this difference is especially apparent for children with high involvement, there would be evidence for the inference that boys with high exposure do view the story as more typical.

In trying to answer these questions a difficulty arises: children with low exposure tend to remember more details in general, and there is no reason to question the generalization of this tendency to more typical details, even if these children do not view them as such. True, in the high involvement group the difference resulting from degree of exposure is not great, and of the children in this group who are highly exposed 77% remember this typical aggressive detail,[21] as opposed to 62% of the equivalent low exposure group. However, there is, perhaps, a better way of approaching this problem.

Because of the fact that Strongman was knocked out, the Communists are able to get on their ship. Strongman revives just as they are leaving the dock and now faces the problem of boarding the vessel. His solution is to hop on a crane on a grain loading gallery which his midget friend and helper operates. This fact—that he boards the boat by means of a crane on a loading gallery—is as important to the plot of the story as is the fact that he was knocked out. However, one cannot say that boarding a ship in this way is an element common to comic book stories in general; rather it is unique to the particular plot of this story. If the hypothesis is correct, children with high exposure ought to remember the typical element—that Strongman was knocked out—more than the unique one,[22] whereas among low exposure boys there should be no such difference. Table 17, which shows the data on this point, indicates that high exposure children are much more likely to remember the typical than the unique detail ($\chi^2 = 16.0$, $P < .0005$), whereas among low exposure boys there is little difference ($\chi^2 = .6$, $.25 < P < .50$).

TABLE 17

NUMBER OF BOYS REMEMBERING TYPICAL AND UNIQUE DETAILS BY EXPOSURE

Unique Detail	Typical Detail: High Exposure Right	High Exposure Wrong	Low Exposure Right	Low Exposure Wrong
Right	13	2	16	10
Wrong	23	12	15	9

Hence, the first premise would appear to be correct. The question still remains whether this difference in type of detail remembered is intensified among children with high involvement.

When memory of the typical element is compared with that of the unique detail of how Strongman got on the boat, no such intensification exists. All children with high exposure remember the aggressive detail more, whether involved or not. The structure of the story is such, however, that these two details are closely related to each other in terms of plot and in terms of spatial proximity. Hence, the highly involved child with high exposure, whose tendency may be to concentrate on this part of the story might, because of this very involvement, show good recall of a unique detail closely

---

[20] See General Procedure in the Introduction for a short description of this story. The comic book itself is available in the dissertation copy (Bailyn, 1956, App. B).

[21] "What happened to Strongman after he upset the fork lift?" If a child did not remember the fork lift incident, the interviewer explained the setting to him in detail to insure that the measure of memory of the aggressive event would not be influenced by memory of a previous event.

[22] "How did Strongman get on the boat?"

TABLE 18

MEAN PERCENTAGE SCORES OF MEMORY OF THE TYPIÇAL DETAIL BY EXPOSURE AND INVOLVEMENT

Exposure	Involved	Uninvolved	Total
High	36.3 ($N = 23$)	31.1 ($N = 22$)	33.8 ($N = 45$)[a]
Low	26.4 ($N = 12$)	25.6 ($N = 34$)	25.8 ($N = 46$)[a]

[a] Nine boys—5 with high and 4 with low exposure—gave no correct answers to the Strongman questions and are excluded from the table.

related to a typical one. Comparison with a unique element from an earlier part of the story[23] shows that among high exposure children, those who are highly involved show a greater differential between memory of the typical aggressive and the unique or neutral detail than do children with less involvement.

In an effort to equalize differences resulting from the place of the neutral detail in the structure of the story, memory of the typical detail was compared with the answers to all other detailed questions about this story, by using the total number of Strongman details remembered as a base and assigning percentage scores.[24] Table 18 shows the mean percentage scores of memory for the typical detail for children with different degrees of exposure and involvement.

This way of looking at the data confirms the conclusion already drawn: children with high exposure show a greater relative recall of the typical detail than do low exposure children. Further, the table indicates that among high exposure boys, those highly involved in the comic book have a greater tendency toward this selective recall than those uninvolved; among children with low exposure, involvement makes little difference. Boys with high exposure and a high level of involvement are especially attuned to the typical aggressive elements in the comic book.

Both the presence of problems in a boy's life and extrapunitive tendencies seem to increase the effect of high exposure on his perception of the comic book. When these two psychological attributes are combined with high exposure the mean percentage score of memory of the typical detail increases to 48.7%. The other three combinations of these two psychological characteristics and exposure show mean scores between 19.0% and 29.9%. This suggests that high exposure attunes especially the child with problems and extrapunitive tendencies to the typical elements of a comic book—makes him view the comic book more as an example of a type, rather than as a unique story.

As an illustration of this point, the free recalls of the story of Strongman by two boys are presented, one with high and one with low exposure. Both boys have many problems; both have high IQ's; both are extrapunitive. Dick, the boy with high exposure, spent eight minutes on the whole comic book, he was highly involved, and remembered a total of nine details. Peter, with low exposure, spent seven minutes on the comic book, was relatively uninvolved, and remembered ten details. When asked to tell the interviewer the story of Strongman, Peter gave the following account:

> "Strongman is trying to show little boys the advantage of leverage. Then he is called out and went to the waterfront. He got into a fight. Someone crept behind him and hit him over the head with a board or pipe. [He] went over [to where] someone asked him to move a meat freezer, I don't know if he moved it or not. He got on the boat—jumped off a grain elevator—and got on the tanker. The FBI said [the boat] had an atom bomb on it to explode in the harbor. Strongman saw men on the deck. He jumped down, banged [their] heads together. [He] stopped the boat and waited for another boat to come and take him off."

This account must be compared with that given by Dick, who differs from Peter in that he is highly exposed. On problems, direction of punitiveness, IQ, and amount of time spent on the comic book the two boys are very similar. Dick's account follows:

> "[It's about] Strongman and a bomb on a ship. Strongman is doing stunts. The telephone [rings]. Somebody wants him to bring a big box into a ship or something. A cop there saw a man with a gun fooling around (not sure [for what]). [He was]

---

[23] "How many cops were necessary to lift the refrigerator?"

[24] Children who did not give the correct answer to the typical question were given a score of 0%; a child who answered only the typical detail correctly received a score of 100%; a child who remembered, for instance, three others besides, received a score of ¼ = 25%.

looking for the box to get something out. The next picture showed Strongman kicking the box over. (Can't remember. [I have] a short memory.) A man on the boat going off told Strongman to be quiet. 'There's that sneak.' [Strongman] chased the man on a big steamer. [They] knocked Strongman out. The little kid told him where the bad men had gone—on the boat. Strongman jumped on the boat. An FBI man said, '[They have] a bomb, [they'll] explode the harbor.' 'Now [they] tell me,' [said Strongman]. He jumps fighting. He crashed through the door to the control room and beat him up. He shoved another man down him. He found the bomb."

Neither child had a clear perception of the plot, which is not surprising in view of the fact that neither one spent more than three minutes on this particular story. Notice, however, the difference in treatment of detail. Peter mentions that Strongman is showing boys "the advantage of leverage" (almost the exact words of the story); Dick reports that Strongman is "doing stunts." Peter states that Strongman was asked to move a "meat freezer" (it was a refrigerator); Dick calls it a "box." According to Peter, Strongman was "hit over the head with a board or pipe" (it was a wooden club of some sort); Dick merely reports that someone "knocked Strongman out." Peter has Strongman getting on the boat from a "grain elevator" (it was actually a grain-loading gallery); Dick's account mentions only that Strongman "jumped on the boat."

Content Areas

In the above discussion the nature of one common element to which high exposure children are particularly attuned has been mentioned: aggressive behavior. What other content areas of comic books are part of the typical pattern as perceived by children with high exposure?

Crime and criminals. One such element is the portrayal of criminals. What leads a person to a life of crime? Under what motivation is he functioning? Comic book stories are seldom concerned with such matters; the villains in them are criminals just because they are bad people, and of course bad people commit crimes. Closely allied with this point is that of the nature of crime itself. Is there a moral question involved or does the problem of crime center on the technical question of whether or not one is successful or caught? The latter point of view is more emphasized in comic books. As a specification, then, of the general hypothesis, one would expect children with high exposure not to be as concerned with the motivational aspects of criminals as low exposure children, and to view the problem of crime more from a technical than from a moral point of view.

TABLE 19

PERCENTAGE BY EXPOSURE OF RESPONSES TO THE QUESTION "WHY DID FRANK BECOME A CRIMINAL?"

Exposure	Concerned Social	Concerned Psychological	Unconcerned	Total
High	36	14	50	100 ($N = 50$)
Low	50	20	30	100 ($N = 50$)

Evidence on these points is obtained from answers the children give to two questions about the first story, which concerns a young criminal, Frank.[25] The children were asked why Frank became a criminal and what mistakes he made. The replies to the former question were separated into those concerned about Frank's motivations and those not. A child, for example, who said that Frank became a criminal because he was "just born bad" was considered unconcerned. Concerned answers were divided into two categories: those centering on social factors—"His mother was working, his father was dead, and he had no job"—and those centering around more psychological factors—"He always wanted his own way; no one gave it to him so he became a criminal to get it." Table 19 shows the percentage of boys' responses in each of these categories.

As can be seen from the table, high exposure boys are less concerned about the motivations behind a life of crime than are children with low exposure ($\chi^2 = 3.4$, $.025 < P < .05$). Of those who give concerned responses, however, boys with high exposure are as likely to give psychological

---

[25] See General Procedure in the Introduction for a short description of this story.

9.3.81
J 739

reasons as are boys with low exposure. This result is surprising: one would expect high exposure children to give fewer concerned responses of a psychological nature.

In this particular story there is mention —in one frame on the first page—of the fact that Frank was brought up on the lower east side of New York and that his mother had difficulty supporting him. Of the low exposure children, 50% mention these facts in their free recall of this story; only 28% of those with high exposure do so. This result probably explains the lack of difference found between exposure groups in the type of concerned reasons given. Even if it were true that high exposure children have a tendency to give fewer psychological reasons, this fact would probably be offset by the tendency of boys with low exposure to be more true to the actual details of the story, which, in this case, center on social factors. This point will be mentioned again later.

A child who said that one of Frank's mistakes was that he "surrendered" and "forgot to learn the law of the state" (Frank did surrender because there was no capital punishment in the state he was in; he had committed a federal offense, however) was considered as giving a technical response to the question of what mistakes Frank made. On the other hand, feeling that Frank's mistake was that "he killed people—everyone likes to live," was considered a moral response. Some children, of course, gave a mixture of these two types; their answers were not classified into either category. Table 20, which gives the percentage of responses falling into these categories by exposure, indicates that high exposure boys give more technical responses to this question than do low exposure boys ($\chi^2 = 3.1, .025 < P < .05$).

TABLE 20

Percentage by Exposure of Responses to the Question "What Mistakes Did Frank Make?"

Exposure	Moral	Technical	Unascertainable	Total
High	24	40	36	100 ($N = 50$)
Low	40	24	36	100 ($N = 50$)

The two boys introduced above provide an illustration for the present point. Dick, with high exposure, feels that Frank made a mistake to "use hand grenades" because he "could have knocked the guy out without killing him" and that "when trying to rob the guy with a knife" his mistake was that he "doesn't know how to use a knife well" or, he added, "a gun for the phone booth man." This response was considered a technical one. His answer to the question why Frank became a criminal—because he "disliked cops"—was classified as unconcerned. He gave no indication, in his free recall of the story, of having noticed the limited background facts given in the comic book.

Peter's list of mistakes—"tied up with the man who gave him a job; trying to make up for his first mistake by trying to kill an FBI man; to kill an officer; breaking out and robbing a bank"—was considered unascertainable, neither clearly moral nor clearly technical. He gave a socially concerned reason for Frank's becoming a criminal: "Didn't have anyone to tell him different, under the influence of other men." His free recall started with the following: "Boy's father died; he got in with the wrong people; he had no one to guide him."

A lack of concern for the motivations of a criminal and a technical, rather than a moral, view of crime, then, seem to comprise another typical aspect to which boys with high exposure are particularly attuned.

Law enforcement. What about the other side of the coin, the view of the recognized law enforcement institutions of society? Some comic books, of course, rely on society's accepted officers of the law as their heroes; others, however, reach into the world of fantasy to supply this need, often to the detriment of the policemen and FBI agents who also appear in the stories. One such story is the last one shown to the children—the story of Strongman. When asked what would have happened if Strongman had not been around, most children answered in terms of the comic book plot: "Gang would have gotten away in the ship" (get away); "Would have set off atomic bomb" (blowing up); "The crooks would have got what they wanted" (generalized success). A few went beyond the story in

TABLE 21

Percentage by Exposure of Responses to the Question "What Do You Think Would Have Happened If Strongman Hadn't Been Around?"

Exposure	Get Away	Blowing Up	Generalized Success	More Threat	Caught Anyway	Total
High	46	30	16	2	6	100 ($N = 50$)
Low	45	31	6	4	14	100 ($N = 49$)[a]

[a] One boy, who answered this question by saying "there wouldn't have been a story," is excluded from the table.

TABLE 22

Percentage by Exposure of Responses to the Question "If the Story Had Continued, What Would Have Happened Next?"

Exposure	Man into Space	Threat to Earth	Other[a]	Total
High	18	34	48	100 ($N = 50$)
Low	28	20	52	100 ($N = 50$)

[a] Of these responses, 46% center solely on the dog in the story: "The master would keep the dog and be more patient with him." Both Peter and Dick gave this type of response. Children who gave a purely dog-oriented response tend to have many problems and be extrapunitive; their exposure is a little more likely to be low than high.

the direction of more future threat: "Would have escaped—left the harbor—and later Communists would have sent destroyers." Only some said that "sooner or later cops, even without help, would have gotten them." As can be seen in Table 21, there is little difference in the answers to this question given by boys with high and those with low exposure. Boys with high exposure give somewhat more "generalized success" responses and somewhat fewer "caught anyway" answers ($\chi^2 = 2.3$, $.05 < P < .10$).

Some children specifically said that the police didn't "dare go near them" or mentioned, in their free recalls, the somewhat snide remark of Strongman when a policeman told him of the bomb in the boat: "*Now* he tells me!" (see, for instance, Dick's recall of this story). Of the boys with high exposure, 26% give indication of having noticed this negative attitude toward the police; 16% of the low exposure boys mentioned it.

There seems, then, to be a slight tendency for children with high exposure to show greater perception of a negative attitude toward the ordinary law enforcement agencies as shown in the third story of the comic book.

Threat to earth. A final content area investigated is that of threat to earth from space. Many space stories, especially those in comic books, center around this theme; hence, the question arises whether the portrayal of such threat is another common element to which high exposure children might be particularly attuned. It should be pointed out, however, that there is another type of space story—most frequent on television—that centers on the adventures of man in space, and though conflicts occur in these stories too, they do not necessarily center on the aspect of threat to earth. The distinction between these two types of space stories was nicely made by one of the boys interviewed when he compared stories like *Space Cadet* (emphasis on man in space)—which he prefers—to the second story in the comic book, which he perceived as a threat to earth type of story. Of the *Space Cadet* type he said "they're not fantastic—they may be crazy but they're not fantastic." Since, however, the present section deals with comic books, which often feature the threat type of story, and since the high exposure children read more comic books than those with low exposure, the point may still be investigated.

One thing asked the children about the space story in the comic book was what they think would have happened next, if the story had continued.[26] One half of the children gave answers not dealing with the problem of earth and space at all, but centering, rather, on the immediate details of the story. Some, however, felt that the people in space would "send more men and conquer earth." A threat response of this nature was contrasted with a response that centered on man going into space: the scientist would have "invented time machine, went there" and would have "made modern things on the moon." Table 22 shows the responses given to this question by children with high and low exposure. It shows a slight tendency for children with high exposure to give somewhat more "threat to earth" responses, those with low exposure to favor the "man to space" type of response ($\chi^2 = 2.0$, $.05 < P < .10$).

---

[26] See General Procedure in the Introduction for a short description of this story.

Characterization

What impression do the children have of the people in the comic book stories? In what terms do they evaluate them? Each child was asked, at the end of his free recall about a story, what every person he mentioned was like. The answers to these questions provide the following data.

Each characterization was classified, first, into three major categories: descriptive, evaluative, or explanatory. Further distinctions can be made within each of these types. Thus, the characterization of a person as "tall, not too skinny or fat; had mustache and black hair" is physically descriptive. A description of the same person as one who "came home from a long trip—had to do with business—was tired and grumpy" is a statement descriptive of action. A third descriptive category is exemplified by the statement "He's the kind of person who would take candy from a baby"—a statement descriptive of character. Hence, descriptive statements were divided into those that were physically descriptive, descriptive of action, or descriptive of character.

Evaluative statements were divided into those in which the terms of evaluation were completely stereotyped—"What was the chief on the ship [Strongman story] like?" "Very bad"—and those in which evaluations were based on specific attributes of the person involved—"[Strongman] didn't brag—he was very nice." There are, then, stereotyped and nonstereotyped evaluative statements.

An explanatory characterization delves into the motivations of a person. "The boy would have been all right if he were brought up right" is explanatory in a social sense; "[The space leader] wanted to be king" is an attempt at explanation in a more psychological vein.

A certain percentage of every child's statements fell into each of these categories. Table 23 gives the mean percentages for all the children. There is no difference between children with high and those with low exposure on the percentage of their statements in each of the major categories: on the average, a little over half of the statements of children with either degree of exposure were descriptive; about one third were evaluative; only around 3% were explanatory. Are there distinctions within these categories?

TABLE 23

Mean Percentage of Characterizations Falling Into Different Categories

Category	Mean Percentages	
Descriptive		54.2
Physical	11.3	
Action	22.7	
Character	20.2	
Evaluative		33.2
Stereotyped	24.0	
Nonstereotyped	9.2	
Explanatory		3.1
Social	2.0	
Psychological	1.1	
No answer, unclassifiable		9.5

Note.—The base on which these percentages were computed—the number of people mentioned—varies greatly from person to person. The mean number for the whole group, for all three stories, is 13.8. Low exposure children mention 14.4 people on the average, high exposure 13.2.

One might recall at this point the following exchange reported by the novelist Steinbeck (1955) of a conversation with his son Catbird, about a cowboy story on television:

> "Now wait," I [Steinbeck] said, "which one is the Good Guy?"
>
> "The one with the white hat."
>
> "Then the one with the black hat is the Bad Guy?"
>
> "Anybody knows that," said Catbird.

To Steinbeck's son, characterization in these stories is obviously portrayed by means of physical attributes and even stereotyped costuming. However, it requires some practice to be able to orient oneself to the structure of a plot in this way. One would expect children with high exposure to be more adept in this way of viewing television or comic book stories.

It turns out that 24.8% of the descriptive statements of boys with high exposure are physical in nature; 20.8% of those with low exposure fall into this category. This difference, though in the expected direction, is not statistically significant. On further investigation, it appears that this type of descriptive characterization seems to be mainly used by high exposure boys who have many problems and who are extrapunitive. Among these children the mean percentage in the physical description category increases to 45.7%; the other combinations show means between 14.4% and 32.4%. This indicates that a child with many problems and extrapunitive tendencies who has high exposure is most likely, in his descriptive characterizations of people, to concentrate on physical attributes.

As can be seen in Table 23, a large percentage of the evaluative characterizations of children are stereotyped in nature, a percentage one would expect to be increased among children with high exposure. Of the evaluative statements of high exposure boys, 70.2% are stereotyped as opposed to 61.3% of those with low exposure. This difference, though in the expected direction, is not statistically significant. Again, however, an intensification of this tendency appears among high exposure children with many problems. In this group, 78.9% of the evaluative statements are stereotyped; the mean percentages in the other groups range from 53.3% to 64.9%.

When extrapunitiveness is added to many problems among children with high exposure, this percentage increases to 81.0%. However, this group does not stand out particularly from that with high exposure, few problems, and intropunitive tendencies (73.4%), or that with low exposure, many problems, and extrapunitive tendencies (77.8%). Rather, in the case of stereotyped evaluative statements, it is the children who have few problems, are intropunitive, and have low exposure who give outstandingly few stereotyped evaluative characterizations: only 52.1% of their evaluative statements are stereotyped.

In discussing the view the children have of the motivations of a criminal the expectation was stated that children with low exposure, if concerned about such motivation, would give more psychological reasons, those with high exposure would give more social ones. At that time this expectation was not realized because of the greater literalness of low exposure children who picked up the social facts given in the story. The present data provide another opportunity to investigate this question. Of the explanatory characterizations given, are those of children with high exposure more frequently put in social terms? Unfortunately, only thirty children—fifteen with high and fifteen with low exposure—gave explanatory characterizations, hence the information is limited. The data on these thirty, however, tend to confirm the hypothesis: of the explanatory statements given by children with high exposure, 82.2% are social in character; only 60.0% of those with low exposure are.

Reality

At the end of the questions on Strongman the children were asked whether they feel that in real life the FBI is frequently helped in this way and whether they think that events such as were portrayed in the story really happen. Children with high exposure were more likely to feel that the FBI is frequently really helped in this way ($\chi^2 = 3.4$, $.025 < P < .05$); a similar, though not statistically significant, trend appeared in the answers to the question directly asking about realism. When asked directly, high exposure and a feeling of realism is associated mainly among children with many problems, as can be seen in Table 24. Children with many problems and high exposure are the ones most likely to view a story, such as Strongman, as in some way representing reality to them.

TABLE 24

PERCENTAGE BY EXPOSURE AND PROBLEMS ANSWERING "YES" TO THE QUESTION "DO THINGS LIKE THIS HAPPEN IN REAL LIFE?"

Exposure	Problems	
	Many	Few
High	60 ($N = 25$)	32 ($N = 25$)
Low	36 ($N = 25$)	32 ($N = 25$)

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....
Accs. No........

Information is available, as has been mentioned before, on parental restriction of exposure time. Further, there is information on whether parents try to restrict children on certain types of comic book stories, of which Strongman is probably an example. Suppose that children are restricted on this type of story, yet read examples of it anyway. Will they then interpret the restriction as keeping a "bad" part of reality from them and consequently view the stories as more realistic? Of the children with low exposure, 34% say "Yes" to the realism question; this percentage changes to 36% among the children whose parents restrict the type, though not the amount, of stories they read. In contrast, among high exposure children the over-all "Yes" percentage of 46% is increased to 68% in the group restricted on type only. Among high exposure boys with many problems, 83% give "Yes" answers when restricted by type only. This intensification of the feeling of realism does not occur among children restricted both on amount and type. These children, perhaps, are used to restriction and put no special interpretation on not being allowed to read a certain type of story.

To summarize, high exposure boys are more attuned than those with low exposure to the typical elements in a comic book: for instance, aggression, threat, an amoral view of crime with little concern for the reasons why a person turns to such a life, a negative attitude toward the police. Further, high exposure children tend to emphasize physical characteristics of the people in the stories, use stereotyped terms in evaluating them, and explain their behavior in terms of social factors. They also seem to have more of a feeling of realism about the story's content. Most of these differences are especially evident among high exposure boys with certain psychological characteristics, mainly problems in their daily lives or extrapunitive tendencies.

Some Cognitive Correlates and Effects of Exposure to the Pictorial Media

In the course of the discussion in the last section, certain relationships between thought processes and high exposure were indicated. The present section centers solely on these relationships: Which of the thought processes under investigation in this study are correlates of exposure, and which of these correlates may be considered effects of exposure?[27]

The four areas of possible correlates under investigation are stereotyping, perception of threat, active or passive tendencies, and the nature of the projected self-image of the child. In the section on psychological and cognitive characteristics, the relation of these factors to social and psychological characteristics was discussed; the present section introduces differences in exposure. Since there are sex differences in these cognitive factors, the discussion treats boys and girls separately.

On the basis of the nature of some of the mass media, especially television programs for children and comic books, the prediction was made that children with high exposure would be more stereotyped in their view of people and would perceive more threat in their environment. These hypotheses are based on the fact that many stories which children see present human beings in black-and-white terms, either very good or very bad, very weak or very strong. Frequently the plot centers on an individual helpless in the face of a threat from external forces, from another country or planet or spy rings within the community; eventually he is helped by some heroic figure, often one with attributes the poor individual cannot possibly have or hope to have. It seems logical to assume that if exposure to the pictorial media has any effect whatsoever, it will consist, partly, of incorporation into the child's thought processes of such repeated elements. The way these stories depict character may lead to stereotyped classification of nonfictional people; the constant reiteration of danger may result in a child's perception of threat in his actual environment. It should be noted that these hypotheses depend not only on the amount of exposure but also on the content of exposure. Not all television programs or comic books contain these elements.

---

[27] The distinction between cognitive correlates and effects was drawn in the introductory section. There it was stated that the experimental data of this study would provide the means by which inferences on effects may be made. The general results of the experiment and the data relevant to the establishment of cognitive correlates are presented in this section, but space limitation makes it impossible to include the details of the experiment. These may be found in the dissertation copy (Bailyn, 1956, Ch. IX). It should be said that technical difficulties in the experimental procedure precluded a definitive test of the hypothesis and the results obtained must be considered tentative.

Stereotyping

Exposure and stereotyping are indeed related among boys: those with high exposure also have higher stereotype scores ($t = 2.3$, $.01 < P < .025$). Of the twelve items, responses to which comprise a boy's stereotype score, four are particularly strongly differentiated by exposure:

(1) Why do people who commit crimes do so?

(2) People are either all good or all bad.

(3) Why do some people tell lies?

(4) There are only two types of people in the world, the weak and the strong.

Boys with high exposure are more likely to answer the first item by the response "Because they are naturally bad" ($\chi^2 = 4.3$, $.0125 < P < .025$), to feel that the second statement is true ($\chi^2 = 2.9$, $.025 < P < .05$), to answer the third by "Because they are naturally dishonest" ($\chi^2 = 3.7$, $.025 < P < .05$), and to feel that the fourth statement is true ($\chi^2 = 3.1$, $.025 < P < .05$). Further, this relationship persists when certain background characteristics (father's occupation, religion, IQ, and amount of parental restriction) associated with both stereotyping and exposure are held constant.

What about the effect of the psychological characteristics on this relation? Is it exposure per se that accounts for this difference in stereotype scores or does an increase in stereotyping appear only when exposure is combined with certain psychological tendencies within the child? Table 25 presents the data relevant to this point with regard to the problems a child has.

TABLE 25

BOYS' MEAN STEREOTYPE SCORES BY EXPOSURE AND PROBLEMS

Exposure	Problems	
	Many	Few
High	4.82 ($N = 40$)	3.94 ($N = 47$)
Low	3.70 ($N = 43$)	3.65 ($N = 46$)

As can be seen from the table, it is mainly the existence of high exposure in conjunction with many problems that leads to a greater tendency toward stereotyped thought. Both the presence of problems in a boy's daily life and high exposure to the pictorial media are necessary to increase stereotyping.

Further, an increase in stereotyping occurs when high exposure is accompanied by extrapunitiveness, as can be seen in Table 26. A boy who is extrapunitive and also highly exposed to the pictorial media will tend more toward stereotyped thought.

Exposure and the combination of many problems and extrapunitiveness produce a still greater degree of stereotyping: children in this group have a mean score of 5.36. In contrast, the scores of the other three combinations range from 3.34 to 3.60.

The third psychological variable in this study — rebellious independence — does not combine with exposure to increase stereotyping in the same way. Both highly exposed and rebelliously independent boys tend to be more stereotyped than those who are less exposed or nonrebellious. In their relation to stereotyping, exposure and independence are additive factors:

	Rebelliously Ind.	*Nonrebellious*
High exposure	4.35(N=31)	4.02(N=41)
Low exposure	3.95(N=20)	3.64(N=47)

The initial hypothesis, however, involves content of exposure as well as amount of exposure. The present study includes information on an indirect measure of content, that represented by the child's preferences. Using this measure as an indicator of the

TABLE 26

BOYS' MEAN STEREOTYPE SCORES BY EXPOSURE AND DIRECTION OF PUNITIVENESS

Exposure	Extrapunitive	Intropunitive
High	4.58 ($N = 38$)	3.83 ($N = 36$)
Low	3.59 ($N = 34$)	3.38 ($N = 34$)

content of exposure one would expect children who prefer material of the aggressive-hero type to be more stereotyped. And, indeed, analysis confirms this expectation. The boys who rate "high" on preference for AH material have higher stereotype scores than those who rate "low" ($t = 2.1$, $.01 < P < .025$).

As might be expected, high exposure to the pictorial media accompanied by preference for AH content was found to increase the boys' tendency to stereotype.

High exposure and preference for AH material are especially strongly related among boys who have many problems and among those who are extrapunitive, as has been indicated in the discussion of boys' exposure habits. Many problems and extrapunitiveness seem to lead a boy to prefer material of the AH variety if his exposure is high; this same child tends to be most stereotyped. Further, there is some evidence that stereotyping is an effect of exposure: low exposure boys with many problems and extrapunitive leanings seem to be most affected by the comic book.

Though the mean stereotype score of girls with high exposure as well as that of girls who prefer AH content is larger than the mean score of those with low exposure and those who have a "low" preference for AH material, neither difference is statistically significant. One cannot say, therefore, that either aspect of exposure differentiates degree of stereotyping among girls.

Perception of Threat

The children's perception of threat is looked at from two points of view: threat to the individual and threat to the country or world. For boys there is no relationship between the former—the personal area—and exposure: neither aspect of exposure is significantly related to scores in the personal area. Nor are these scores increased in a marked way among boys with high exposure and a preference for AH content. It seems that boys' feelings about the relative helplessness of the individual are not related to exposure to the pictorial media. The same is true for the girls.[28]

Nor among boys does exposure make a difference in the perception of threat to the world. In the case of girls, however, a different picture emerges on this point. Girls who are highly exposed to the pictorial media perceive somewhat more threat to their country or world than do those with low exposure ($t = 3.1$, $.0005 < P < .005$).[29] One item is particularly strongly differentiated by exposure: Do you think there is danger that people on this earth will be harmed by beings from another planet? Girls with high exposure are much more likely to think such an occurrence possible than are those less exposed ($\chi^2 = 10.0$, $.0005 < P < .0025$). Surprisingly, preference for AH material does not differentiate girls' perception of threat to their country or world. If anything, girls who prefer this type of content give fewer threat responses than girls who do not. And, as a matter of fact, the girls with high exposure and no particular preference for AH content have the highest scores in the world area.

Passivity

The hypothesis being tested here is that exposure to the mass media is related to passivity. This view is based on the presumption that a high amount of exposure, which represents a passive form of behavior, may be associated with a child's preference for passivity in general, a lack of initiative, and a receptive attitude toward the things around him. Two sets of data are relevant to this question. The first concerns the preference given by children, in a two-choice situation, for a more active or a more passive form of social behavior.

---

[28] This lack of correlation might be explained by the following finding of a study about children and movies (Shuttleworth & May, 1933). When presented with a story in which the hero is threatened by extreme danger and asked to predict what will happen, children who seldom attend movies tend to feel that the hero will succumb; children who see many movies, on the other hand, are more likely to feel that he will escape, even when the means of this escape are highly improbable. Thus the effect of constant reiteration of threat might be offset, among children with high exposure, by the belief in the certainty of escape.

[29] Some of this relationship, however, is accounted for by parental restriction: among restricted girls the relation disappears; among unrestricted ones it is reduced.

Boys with low exposure to the pictorial media give more active preferences than do boys with high exposure ($t = 2.4$, $P = .01$). Of the five items used to determine active preference, two concern a choice between exposure to a medium and an alternative more active activity. In the other three items the active choice represents initiative as opposed to a more receptive attitude implied in the passive choice. Between going to a movie or on a picnic high exposure boys choose a movie more often ($\chi^2 = 4.7$, $.0125 < P < .025$), and much more frequently pick watching television over playing a game indoors ($\chi^2 = 14.5$, $P < .0005$) than do low exposure boys. However, responses to the other three items, which concern starting a new club or being asked to join one, giving a party or going to someone else's, deciding which games to play or letting others decide, are not differentiated by exposure. In other words, even though high exposure boys prefer exposure to the mass media to more active alternative activities, this preference for passivity does not generalize to a lack of initiative and a preference for being on the receiving end of a social situation. Hence, on the basis of these data, the hypothesis, though seemingly statistically confirmed, reduces almost to a tautology: boys with high exposure enjoy exposure to the media.

There is, however, another set of data that will throw light on this question. As has been indicated, each child was asked what occupation he would like to have when grown up. These choices were then classified according to whether or not, in view of the child's capabilities and background, they were realistic. Further, the realistic choices were placed into two categories, representing either passive acceptance by the child of the occupational status of the father or active attempts to change this status in the projected future. Among boys, this classification was made on the basis of a comparison of the child's choice and the occupation of his father. Hence, the hypothesis may be restated in terms of this material: boys with high exposure should give fewer active occupational choices than boys with low exposure. It turns out that among high exposure boys an average of 24% of their realistic choices are active in this sense. The mean for boys with low exposure is 40%, a significant difference ($t = 2.2$, $.01 < P < .025$). In regard to occupational ambitions, then, high exposure among boys does seem to be associated with a more passive attitude than does low exposure.[30]

When high exposure is accompanied by many problems, this tendency is intensified: only 19%, on the average, of this group's choices are active—represent a hoped-for occupational status different from that of the father.

This hypothesis is not confirmed for girls. Exposure does not affect the number of active preferences given by the girls; neither does it affect their number of active realistic occupational choices.[31]

Projected Self-Image

The last area to be discussed concerns the projected self-image of the child: what type of person he would like to be or the type of work he would like to do. The general hypothesis here is that a child's self-image is colored by exposure to the pictorial media. The responses to two questions provide the data on the basis of which this hypothesis is investigated: the type of occupation the child wants to have when grown up, and the persons, from a list of seventeen, he would most want to be changed into. The former represents a child's projected self-image in a realistic context, the latter in a fantasy context.

---

[30] Clearly, a child whose father is a top professional has less opportunity to have his occupational choice scored as active than a child whose father is a worker, for instance. Because of this ceiling effect the results obtained give a minimum measure of the difference. It will be recalled that the fathers of boys with high exposure are more likely to be workers than are those whose sons are less exposed. Hence, the ceiling effect works against the hypothesis.

[31] Only on Item 22—television or game—do girls with high exposure more frequently choose the passive choice of television: $\chi^2 = 4.6$, $.0125 < P < .025$. Not even the responses to the other item concerning exposure to a medium are differentiated by exposure among girls.

Realistic. The three categories into which most of the boys' first occupational choices fall, as has been seen earlier, are top professional, athletics, and worker. Boys with high exposure introduce a new category which ties with athletics as the third most frequently chosen one. This category might be labeled "cops and robbers"; it includes occupations like policeman, detective, FBI agent. Of the choices of boys with high exposure, 11% fall into this category as opposed to only 4% of the choices of the low exposure boys. This difference, though reduced in some cases, is not obliterated when relevant social characteristics are held constant.

Among boys who, besides being highly exposed, also have many problems, "cops and robbers" proves to be the modal category. More first choices of boys in this group fall into this category than into any other. In contrast, the modal category of the other three exposure-problem groups is top professional or, in the case of the group low on both, engineer.

Among girls there is a similar phenomenon with respect to the category of entertainment. Occupational choices like actress and singer occur more frequently among high exposure girls than they do among girls with low exposure. In the group of girls with high exposure and few problems this category is the modal one. All the other exposure-problem groups have lesser professional in general or teacher specifically—the two most frequently chosen categories of the girls as a whole—as their modal category. This result lends support to a conclusion already drawn: it is only the girls with few problems who seem to show, at times, some deviation from the fairly uniform norms of girls' behavior.

Exposure, therefore, is not unrelated to the nature of the child's occupational choices. But can one say that occupational choices of children with high exposure are derived from the content of the pictorial media? A difficulty is immediately apparent. Take, as an example, a boy who wants to be a policeman. One's first conclusion might be that this choice is derived from the media. His father, however, might be a truck driver who has a policeman as a very good friend; this direct contact might be the source of the boy's choice. Whether such an influence does or does not exist in the child's life one cannot, from the present data, know. The best that one can do is to infer the probability of such an event. One can say that the probability of such direct contact and resulting influence is higher in the life of a truck driver's son than it is in that of a boy whose father is an architect, for instance.

Classifying occupational choices as realistic or unrealistic takes account of this consideration. The truck driver's son's choice of policeman would be classified as realistic, that of the architect's son as unrealistic. Hence, the question of media orientation of occupational desires is best investigated by using only unrealistic choices. All such choices were classified according to whether or not they are media oriented—whether they seem to be derived from occupations frequently portrayed in the pictorial media.[32] It was expected that children with high exposure would have a larger percentage of their unrealistic occupational choices oriented in this way. The data, however, do not confirm the hypothesis.

In the over-all realistic context of this question, there is no relation between the child's projected self-image and his exposure to the pictorial media. The expectation that this self-image is colored by exposure is not realized in this context. But what happens when the context is changed to a clearly fanciful one?

Fantasy. As indicated earlier, the modal personality that boys want to be changed into is Superman, followed in order by animal breeder and FBI agent. Table 27 shows the percentage of boys' first choices falling into each of these categories for each exposure group.

---

[32] Each child was given a score based on the percentage of his unrealistic choices—of those that were classifiable in this respect—that were media oriented. See Appendix C for the exact coding scheme.

TABLE 27

PERCENTAGE OF BOYS' FANTASY METAMORPHOSES IN TOP THREE CATEGORIES BY EXPOSURE

Category	Exposure	
	High	Low
Superman	21	9
Animal breeder	10	15
FBI agent	13	7

As can be seen from the table, boys with different amounts of exposure have very different metamorphic desires. As a matter of fact, the top three categories of boys with high exposure—Superman, FBI agent, detective—are completely different from those of the low exposure boys, which are animal breeder, president, and inventor. The girls' favorite personalities—movie star, princess, animal breeder—do not vary between exposure groups except in degree of choice: 54% of the girls with high exposure give movie star as their first choice as opposed to 39% of the girls with low exposure.[33]

Of the seventeen personalities from which the children could choose, eight are considered media oriented: cowboy, movie star, detective, reporter, Superman, space cadet, FBI agent, and person from Mars. On the basis of the number of choices that fall into this group, each child was given a score on degree of media orientation of his fantasy metamorphoses.[34] It turns out that boys with high exposure have a higher media orientation score than do those who are less exposed ($t = 3.2$, $.0005 < P < .005$). Among girls, the difference is in the same direction but is not statistically significant.

How do the psychological characteristics combine with exposure in this respect? Among the boys, all three of the psychological variables are related to degree of media orientation of fantasy metamorphoses: boys with many problems give more media-oriented choices, as do those who are extrapunitive, and especially those with rebelliously independent tendencies. These factors combine with exposure in an additive manner producing highest media-orientation scores among boys with high exposure who are rebelliously independent, extrapunitive, and have many problems—in that order. No such relationships appear for the girls.

It seems, therefore, that the child's self-image is colored by exposure to the pictorial media but only in a context of fantasy. This effect is most marked among boys who are rebelliously independent. Further, the experimental data give some indication that rebelliously independent boys with low exposure are most affected by the comic book, hence lend some support to the conclusion that this cognitive correlate may be considered an effect. There seems to be no carry-over of this influence into the area of more realistic self-appraisal as seen in the children's occupational desires.

SUMMARY AND CONCLUSIONS

This study has investigated the exposure habits of a selected group of children and the influences of the mass media on their thought processes. From an analysis of the extent of exposure of children two distinct clusters of media emerged. Radio has taken on a new role, most likely because of the advent of television. It now forms a unit with books. The pictorial media—movies, comic books, and television—stand in a separate group. Children who listen to the radio and read books tend not to be highly exposed to the pictorial media and, conversely, those heavily exposed to the pictorial media do not spend much time listening to the radio or reading books. Girls are somewhat more often radio and book fans; boys concentrate more on the pictorial media.

The main part of the analysis centered on one of these clusters: the pictorial media.

---

[33] There is some reduction in this difference when the social characteristics are held constant but it does not disappear or reverse. Among boys, too, these differences persist when the social characteristics that relate to these choices are held constant, though in some subgroups to a lesser extent.

[34] A first choice that was media oriented received 3 points, a second 2, and a third 1. Hence, the media scores of children that gave three choices range from 0 to 6. Animal breeder was considered unascertainable in this respect; it was neither considered media oriented nor not so oriented. Only children who gave three ascertainable choices are included in the analysis.

Four factors seem related to the amount of exposure to the pictorial media. These are, in the order of their importance, parental restriction on amount of exposure, the child's IQ, occupation of the father, and religion. Children whose parents attempt to restrict their exposure time actually see less of the pictorial media; children with high IQ's spend less time in this form of entertainment, as do children whose fathers have white-collar occupations and who are Protestant. Hence, the factors associated with high exposure to the pictorial media are lack of parental restriction, low IQ, fathers with worker or service occupations, and being a Catholic. These factors are independently related to amount of exposure.

The next concern was the type of programs preferred by the children. A type of story that centers on a hero aggressively defending himself and others, as, for instance, in the Superman type of story, in war and spy stories, in westerns, crime stories, and space stories, was isolated for particular analysis. This type of material was referred to as the aggressive-hero type, or AH content. There is a large sex difference in preference for this type of material: girls show very little liking for AH content, boys show much more. Four factors are important in explaining a boy's liking for this material. Certain practices of the parents toward their children play a role: children whose parents restrict the amount of exposure time show less preference for this type of material; children who are spanked show a greater preference. Another determinant is rebellious independence. A boy who shows rebelliously independent tendencies will also show a greater preference for AH content. Finally, the child's IQ plays a role here too: the boy with high IQ is less likely to prefer this type of material. The factors associated with a particular liking of AH content, then, are lack of restriction on exposure time, spanking by parents, rebellious independence, and a low IQ. Further, boys with a great deal of exposure are more apt to like this type of material than are those with low exposure. This association is particularly strong among boys with many problems about themselves, their friends, and their families; extrapunitive leanings; or rebelliously independent tendencies.

Among girls there is little variation in the over-all lack of preference for this type of material.

The next area of investigation centered on the reaction of boys to a specific comic book of the aggressive-hero type. High exposure boys, especially when they have many problems and extrapunitive tendencies, see such a comic book more as a type than as a unique story. They are attuned to the typical elements in it; because of the character of the genre, these typical elements stress aggression, threat, a technical view of crime with a lack of concern for a criminal's motivations, and a negative attitude toward law enforcement officers. These boys, with high exposure, especially when they have many problems, view the content of such stories as more realistic than do the others. In their way of looking at the people in these stories, their emphasis is on physical characteristics, stereotyped evaluations, and explanations centering on social factors.

The final set of results centered on the cognitive correlates of exposure—the thought processes that seem to be related to exposure to the pictorial media. Again the first observation was the difference between boys and girls. Whereas boys do seem to show some influences, girls show hardly any. Three cognitive correlates of high exposure were isolated among boys. First, boys with high exposure tend to classify people in a stereotyped way—into black-and-white categories. For them, the reasons for a person's falling into the black category, for instance, are magical; boys with high exposure show little concern for the problems of motivation. This tendency is especially apparent among boys with many problems and extrapunitive leanings. Further, there is some evidence that among boys with these psychological characteristics exposure increases the tendency to think in this way. Second, high exposure is related, among boys, to a passive attitude—an attitude of acceptance of the socioeconomic situation of the father and a lack of interest in changing it. This attitude is especially strongly associated with exposure among boys with many problems. Finally, the fancied self-image of boys with high exposure is colored by the mass media. The people chosen by these boys as possible candidates for a metamorphosis—people they would like to be changed into—are taken from the content of the pictorial media. This media orientation is most evident among boys with rebelliously independent tendencies, and some evidence exists that it is an effect.

Among girls, no cognitive correlates were found. There is a tendency for girls with high exposure to feel more threat to the world; otherwise cognitive differences that exist among them are not associated with differences in exposure to the pictorial media.

The Function of the Mass Media

The problem of this study centers on the effect of the act of exposure on a child's cognitive processes. But, as suggested in the introduction, in order to understand even such a limited influence it is necessary to take into account a number of intervening influences. The elements that intervene between the mass media and their effects seem scattered and difficult to relate. There is, however, one conception that unites them: the concept of function. The function of the mass media in the life of the child reflects his psychological predispositions; it also influences his modes of perception. Hence, if one understands the function of the mass media in a child's life, one has progressed materially toward understanding the process of effects.

The main question of this section, then, is whether the results of this study provide empirical support for the existence of a particular function of the mass media in the life of the child. Many functions have been suggested in the discussions of mass media and their effects. Do the present data throw light on this theoretical discussion? We believe they do. Our results support the hypothesis that for certain children, under certain conditions, the mass media serve one function in particular—that of escape.

The Function of Escape

One may conceive of the mass media as providing a substitute world into which one may escape from the problems and tensions of life. Energy resulting from tension within the child is directed toward temporary solutions of problems, solutions having no bearing on the problems as such. Escape into such temporary solutions takes on great importance when, in becoming habitual, it tends to preclude more realistic and lasting solutions. Furthermore, the person for whom the mass media provide a habitual form of escape may so orient himself to the world of the media—a world most likely very different from his own—that he will incorporate the values and behavior patterns of this world. The degree of resulting harm will depend, of course, on the type of content chosen to fulfill this purpose, but the very fact that this media world is disassociated from the one in which the person who exposes himself lives increases the probability that behavior based on such incorporation will be inappropriate and hence problematic.

Opposed to the point of view that the function of the mass media is to supply a substitute world into which one can escape, is the view that the mass media serve functions that may be classified under the heading of *supplementation.* The term *supplementation* is used to highlight the bringing together of the mass media with a person's everyday life, in contrast to the separation of these spheres implied by the function of escape. According to these views, the mass media widen the horizons of the person exposed to them, supply him with information on how to behave in certain situations, or merely offer him needed relaxation. Whether or not there will be a positive influence of the mass media on the person using them in a supplementary way depends upon the congruence of the content chosen and the particular situations in which he finds himself. However, the notions of supplementation, based, as they are, on the relation of the media world to one's own, should minimize any negative effects that might exist.

The results of the present study indicate that high exposure serves the function of escape for some children and not for others. It is the presence of certain psychological characteristics in the child that seems to determine these differences. Specifically, it appears that a child with high exposure who has problems in his daily life, has extrapunitive tendencies, is rebelliously independent will view the mass media from the substitute point of view—use them as an escape. For the high exposure child *without* these psychological characteristics, the mass media do not appear to serve this function.

Substantiation of the view that these psychological characteristics determine whether or not the mass media are used as an escape, rests on two types of arguments: First, a theoretical analysis of their meaning in relation to the logical requirements implied in the function of escape; second, corroboration of these interpretations by means of empirical results of this study.[35]

Logical considerations. It will be recalled that three psychological characteristics have been used in this study: presence of problems, extrapunitiveness, and rebellious independence. The first is essentially a measure of the child's ability to relate to others in his immediate environment, a fact explicit in the peer and family areas. But even the self area fits into this picture. When a child says he wants to be smarter than he is, he certainly is not basing this judgment on an absolute scale of intelligence, but rather on a relative one—relative to the intelligence of the people around him and the situations that confront him. Being in the high problem group, then, is indicative of a child's inability to relate successfully with those around him.

The extrapunitive child is one who directs blame outward. A situation arises with potential ill effects; the blame is not his, it is that of an external agent. If such a child is faced with a difficulty, he most likely blames others for this situation. Hence, we might expect this child to search for solutions to any problems he may have in external sources, rather than within himself.

Rebellious independence, it will be remembered, is measured by the child's unwillingness to accept parents or siblings as companions in certain circumstances and his willingness to proceed alone. A boy with this characteristic is more adventure oriented—he will frequently want to go into the armed services or have an air or sea adventure occupation; he seems oriented to experiences distant from those to which he is accustomed.

These psychological characteristics are seen, then, to have a common element: their presence seems to indicate certain difficulties in a child's life and an accompanying orientation away from the immediate situation in which he finds himself. In the discussion of the meaning of the function of escape these elements were seen to be crucial. This function presupposes, first, a desire to escape. The above analysis has shown that this requirement is met. Further, however, the escape function implies the use of the mass media to fulfill this desire—to provide for these children an alternative point of orientation. In turning to this second requirement one shifts from logical considerations to empirical data for corroboration.

Empirical considerations. The preference for AH content provides a starting point. This content centers on an environment remote from that of the children studied. Such content—as opposed to the *I Love Lucy's* of the mass media—is congruent with the notion that the media fulfill a child's desire to escape. Hence, one can obtain support for the hypothesis by looking at the content preferences of high exposure boys with and without these psychological characteristics.

One of these characteristics—rebellious independence—is related to preference for AH content among all boys. Among high exposure boys, those who are rebelliously independent show particular preference for AH content. Similarly, the high exposure boy with many problems is much more likely to prefer AH content than is the one with few problems. The same holds true for high exposure boys who are extrapunitive. The presence of this psychological characteristic, too, increases preference for

---

[35] These corroborating data are not the results for which an interpretation is being sought in this section. They merely help to establish the general scheme by which these main results—the influence of the mass media on certain cognitive processes of the child—may better be understood. Because of the lack of cognitive correlates of high exposure among girls, the empirical results discussed in this section are mainly relevant to boys.

AH content. Boys with high exposure who have these psychological characteristics show preference, then, for a type of material in the mass media that could well serve the function of escape for them.

There is also another type of supporting evidence: that stemming from the reactions of boys to a particular example of this type of material—the comic book shown the boys in the experimental subsample. Highly exposed boys with these psychological characteristics are particularly likely to view the comic book story as typical, to remember mainly the typical details. The people in the stories, if reacted to on a level beyond mere appearance, are simply part of the black-and-white nature of the media world. Further, such a child views the content of the stories as more realistic.

These differences of reaction indicate that elements of the content these children prefer are organized by them into a fairly stable structure into which a new example is assimilated. Such a structure is a *sine qua non* if the content of the media is to be used as a substitute world of orientation, a notion implicit in the function of escape. It is the presence of such an organization of the material viewed that mainly differentiates between people using the media as escape and those using them, for instance, as relaxation. The latter may like the same type of material, but will not find it necessary to make a meaningful whole of it, since they do not depend on it as a point of orientation. Hence, there is evidence that the media serve an escape function for the boy with many problems, extrapunitive leanings, and rebelliously independent tendencies.

Cognitive Correlates and Escape

In the last section it was shown that a boy with high exposure, many problems, and extrapunitive leanings, shows a tendency toward stereotyping. This attitude is a direct reflection of the content of the mass media preferred by such a boy, elements of which content he has organized in his mind into a more or less stable structure. This child then is asked about his view of people in general, not people from the mass media. On what can he draw for his attitudes? The problems in his own life have not led to speculation about himself, since these problems are not his fault; he can do nothing about them. Hence, his main source of judgment lies in the world he has built up, a world centering around the elements of mass-media stories. And his attitudes will be put in these terms: people in general will fall into black-and-white categories, fiat rather than reason will rule—a fiat stemming from the media world reflected in his imagination.

The second cognitive correlate of high exposure among boys is the impingement of the mass media on their fancied self-image. This tendency is particularly strong among children who are rebelliously independent, though extrapunitive tendencies and many problems also add to it. This child, as has been indicated, adds adventure orientation to the characteristics already described. The world of the media provides an anchorage for such a child's adventure-seeking solutions of his present difficulties. Hence, when asked whom he wants to be changed into—asked to put himself into an imaginary alternative role—he relies on this world to supply an answer.

Third, passivity is related to high exposure, particularly among boys with many problems. Here is a boy with difficulties, using the media as an escape from them. He channels his dissatisfaction into mass-media behavior, providing temporary solutions only. Once, however, his energy has gone in this direction he can stop his search for solutions. An alternative channel to which he might turn—that of change in his adult stage by planning now on realistic occupations different from those in the context of which his family functions—is therefore not utilized by him.

Thus, if a child has certain psychological characteristics he uses the media as an escape and they will have an effect on his thought processes. If he does not have the relevant psychological characteristics—if he does not have many problems, does not have extrapunitive tendencies, is not rebelliously independent—the mass media will not serve as an escape, and will not have these effects. Whether the time of such a boy could be better spent in other activities is another question, which will not be treated here.

Underlying all these considerations of cognitive effects is the basic fact of exposure. The determinants of exposure, as has been indicated, are, on the whole, social. Occupation of father, religion, IQ, and parental restriction affect the amount of a child's exposure. The psychological characteristics of children do not differentiate between those with high and those with low exposure. For the process under discussion, exposure is a prerequisite. Taking this fact into account, the argument may be summarized in the following way.

The amount that a child is exposed to the pictorial media is determined mainly by the social environment in which he finds himself as well as by his general intelligence. Given, however, a pattern of high exposure, the child's psychological characteristics become very important. They determine the function of the mass media for him; they differentiate the type of content he prefers and how he perceives it; they determine the influences of this content on him. A boy with problems, extrapunitive leanings, and rebelliously independent tendencies who, mainly as a result of his social environment and his IQ, is highly exposed, relies on the media for temporary solutions to difficulties; structures the content of the media in black-and-white terms around elements of aggression, threat, amoral views of crime and negative attitudes to existing law enforcement institutions; and draws on this structure in his judgments of people and in a projection of his own self-image.

Proportion of population involved. One can see, therefore, the emergence of a syndrome out of the conjunction of high exposure to the pictorial media and certain psychological characteristics. The value put on this syndrome—whether it is considered harmful or not—is a matter for individual decision. To this question the results of this study provide no answer. One may, however, profitably point out the percentage of boys in a community such as has been studied here that is likely to be involved. In this study only 3% of the boys are highly exposed, have many problems, are extrapunitive, and are rebelliously independent. It has been shown, however, that not all three psychological characteristics are necessary to produce the syndrome. When these elements are taken singly, the percentages rise: high exposure and many problems are combined in 12% of the boys studied; 11% are highly exposed and extrapunitive; 9% combine rebellious independence with high exposure.[36] Nevertheless, it seems clear that the group of boys with the relevant characteristics is not large.

Mass Media and Overt Behavior

One may now be in a position to diverge somewhat and speculate on the implication of these conclusions for the problem of the relation between the mass media and overt behavior. This study indicates that a child with certain psychological characteristics who is highly exposed, picks up certain elements of the mass media, forms these into a picture in his mind, and draws on this picture when asked to express attitudes and desires on related matters. One might speculate, therefore, that in a situation requiring some overt act, which is similar in certain crucial respects to some media-determined picture in his mind, he will also draw on this picture to guide his actions. One would, therefore, expect the mass media to have an effect on the overt behavior of a high exposure boy with certain psychological characteristics who finds himself in a situation to which the structure built up in his mind from the content of the media is applicable. Hence, one may say that the danger area with regard to overt behavior lies in the group of boys with a certain combination of social and psychological factors, a group seen, at least in the community of the present study, to be small. It is in this group that the media serve the function of escape and that the cognitive syndrome investigated makes its appearance. Whether a translation into overt behavior occurs will depend on the situation the boy finds himself in and, possibly, on the amount of tolerance he has for unrealistic solutions. At any rate, one may safely conclude that any connection between the mass media and overt behavior will be indirect, mediated by all the factors discussed.

---

[36] In evaluating these figures it must be remembered that the cutting points of these distributions—with the exception of that of rebellious independence—were empirically determined with a view toward equalizing "highs" and "lows."

It is of interest to note, in this connection, that traits similar to the psychological characteristics of this study were found to play a role in the "causal complex" of delinquency (Glueck & Glueck, 1950). "A far higher proportion of the delinquents display emotional tension arising from a faulty father-son, mother-son, or sibling-boy relationship . . . or out of faulty relations with companions . . . also, there is a higher incidence among the delinquents of conflict-response to feelings of inferiority . . . " (p. 252)—many problems. "A larger proportion of the former [delinquents] give vent to conflict by extroversion of action . . . [which] 'implies that a boy refuses to take any responsibility for his behavior' " (p. 251)—extrapunitiveness. Delinquents "are considerably less cooperative with and dependent upon others" (p. 241); and are "inclined more to adventurous activities" (p. 167)—rebellious independence. The children that the Gluecks studied all came from underprivileged areas; hence, social characteristics were the same for delinquents and nondelinquents.

Effect of the Mass Media on Girls

The most important point to be made in regard to girls is that in the terms of this discussion the mass media have little effect on them. There are really no cognitive correlates of exposure to the media among girls. As a result, the interest shifts in the case of the girls to the problem of explaining the great uniformity of their behavior, their lack of deviation from a norm. Certain parental practices were found to be more uniform when applied to girls than when applied to boys. Also, girls seem to be more adjusted to the school situation—they get much higher conduct and effort grades; they try harder in school—they are much more likely to be over-achievers. In some way, then, they are more conforming and are rewarded for this conformity.

Now why is it that the mass media have less differentiating influence on the girls? First, two of the psychological characteristics important for such an influence—extrapunitiveness and rebellious independence—do not occur very frequently among the girls studied. But girls do have problems. Why does not this factor lead to influences? The answer probably lies in the phenomenon of uniformity itself. Most likely this uniformity reflects a need to conform within the girl herself, a need which is particularly important for girls with many problems. Such girls may seek the solutions of their difficulties by emphasizing an already existing tendency toward conformity. In such an effort help is not supplied by the mass media, except, perhaps, in the refusal to deviate from the normal habits relating to them. And, as has been indicated, any deviation that does exist in relation to the mass media, is particularly evident among girls with few problems—girls whose need to conform is somewhat less.

One gets a picture, then, of girls trying to fit into an established pattern. They expose themselves to the mass media when it is accepted. They will watch television and go to movies, but comic books they read much less—this type of exposure is not usually part of the family pattern. Rather, they emphasize books. They show very little preference for AH content. As opposed to the boys, they emphasize situational stories more than animal ones, a preference pattern most likely shared by their parents. The mass media do not seem to supply a solution to difficulties in a context of conformity to standards of the immediate environment—a context applicable to the girls studied.[37]

The material presented in this study has served a double purpose: to investigate the exposure habits of children to the mass media and to reach some conclusion as to the effect of this experience on certain of the child's cognitive functions. Such an effect seems to exist when exposure is accompanied by certain psychological characteristics in the children themselves.

---

[37] The age under consideration may be very important in this respect. It is entirely possible that at an older age, with content centering around romance and love instead of on an aggressive hero, the media do serve as an escape for certain girls, and do have specific influences on them.

BIBLIOGRAPHY

Albert, R. S. The role of mass media and the effect of aggressive film content upon children's aggressive responses and identification choices. *Genet. Psychol. Monogr.,* 1957, **55,** 221–285.

Bailyn, Lotte. Mass media and children: A study of exposure habits and cognitive effects. Unpublished doctoral dissertation, Radcliffe Coll., 1956.

Butterworth, R. F., & Thompson, G. G. Factors related to age-grade trends and sex differences in children's preferences for comic books. *J. genet. Psychol.,* 1951, **78,** 71–96.

Charters, W. W. *Motion pictures and youth, a summary.* New York: Macmillan, 1933.

Dale, E. *Children's attendance at motion pictures.* New York: Macmillan, 1935.

Glueck, S., & Glueck, Eleanor. *Unraveling juvenile delinquency.* New York: Commonwealth Fund, 1950.

Holaday, P. W., & Stoddard, G. D. *Getting ideas from the movies.* New York: Macmillan, 1933.

Klapper, J. *The effects of the mass media.* New York: Bureau of Applied Social Research, 1950. (Mimeographed)

Lazarsfeld, P. F., & Rosenberg, M. (Eds.). *The language of social research.* Glencoe, Ill.: Free Press, 1955.

Lind, Katherine N. The social psychology of children's reading. *Amer. J. Sociol.,* 1936, **41,** 454–469.

Maccoby, Eleanor E. Why do children watch television? *Publ. Opin. Quart.,* 1954, **18,** 239–244.

Maccoby, Eleanor E., Levin, H., & Selya, B. The effects of emotional arousal on the retention of film content: A failure to replicate. *J. abnorm. soc. Psychol.,* 1956, **53,** 373–374.

Mahony, Katherine St. J. Elementary school pupils' TV habits and choices. *Cath. educ. Rev.,* 1953, **51,** 234–245.

Mitchell, Alice M. *Children and movies.* Chicago: Univer. of Chicago Press, 1929.

Peterson, Ruth C., & Thurstone, L. L. *Motion pictures and the social attitudes of children.* New York: Macmillan, 1933.

Portis, B. An empirical analysis of Catholics' choice between public and parochial primary schools. Unpublished honors thesis, Dept. of Social Relations, Harvard Univer., 1956.

Ricciuti, E. A. Children and radio: A study of listeners and non-listeners to various types of radio programs in terms of selected ability, attitude, and behavior measures. *Genet. Psychol. Monogr.,* 1951, **44,** 69–143.

Riley, Matilda W., & Riley, J. W., Jr. A sociological approach to communications research. *Publ. Opin. Quart.,* 1951, **15,** 445–460.

Rogers, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contr. Educ.,* 1931, No. 458.

Rosenzweig, S., Fleming, Edith E., & Rosenzweig, Louise. The children's form of the Rosenzweig picture-frustration study. *J. Psychol.,* 1948, **26,** 141–191.

Shuttleworth, F. K., & May, M. A. *The social conduct and attitudes of movie fans.* New York: Macmillan, 1933.

Steinbeck, J. How to tell good guys from bad guys. *The Reporter,* 1955, **12** (5), 42–44.

Strang, Ruth. Why children read the comics. *Elem. Sch. J.,* 1943, **43,** 336–342.

Witty, P. Children's interest in comics, radio, motion pictures, and TV. *Educ. Admin. Supervis.,* 1952, **38,** 138–147.

Wolf, Katherine M., & Fiske, Marjorie. The children talk about comics. In P. F. Lazarsfeld & F. N. Stanton (Eds.), *Communications research, 1948–49.* New York: Harper, 1949. Pp. 3–50.

(Accepted for publication February 5, 1958)

APPENDIX A

COMMUNITY AND SAMPLE

Community. The town of Arlington, Massachusetts, in which this study was made, has a population of slightly over 47,000. It is a suburban town located a few miles north of Boston. It contains nine public and two parochial elementary schools. Almost 30% of the fifth and sixth grade school population were enrolled in parochial schools at the time of this study.

Arlington as a whole seems fairly heavily Catholic. Judging from enrollment figures in the parochial and public schools and from the number of Catholics in the present sample, it seems as if almost two thirds of the fifth and sixth grade population are Catholic, of which slightly over half are enrolled in public schools. This portion is, therefore, a selected one. A study of Catholics' choice of parochial or public schools conducted in Arlington (Portis, 1956) shows that white-collar families are more likely to send children to parochial schools, as are families in which neither parent went to college. This study also shows that Catholics are more likely to send daughters than sons to parochial schools.

Sample. The sample used in this study consists of the fifth and sixth grade children in seven of the nine public schools of Arlington. One school had burned down in the spring of 1954, and a second close to it was on double session taking care of both schools, hence these two schools could not be used.

The questionnaire was administered during the week of March 14th, 1955, to every fifth and sixth grade class in the seven schools—a total of 629 children. The final sample consists of 626 of these children, 332 boys and 294 girls. Three cases were dropped: one boy filled in only a few pages of the questionnaire because he was at such a low level of reading that he could not handle more; two other children (one boy, one girl) left too many questions unanswered to be useful for the analysis.

Experimental subsample. The boys in the experimental subsample were selected from among the four extreme exposure-problem groups: high exposure-high problems (H-H); high exposure-low problems (H-L); low exposure-high problems (L-H); low exposure-low problems (L-L).

Of the 179 boys in these extreme groups, 100 were selected, 25 for each of four experimental groups: H-H, H-L, L-H, L-L. The selection procedure was as follows: First, all boys whose frequency of reading comic books was "1 every few weeks" or less were eliminated from the high exposure groups; similarly, those who said they read "2 or 3 a day" or more were eliminated from the low exposure groups, in order to insure definite differences between high and low exposure groups on frequency of exposure to the medium used in the experimental part of the study. Second, the boys whose reading grades in the year 1954–55 contained a D or F, or who were on a reading level of a lower grade than the one they were in, were eliminated, since one of the tasks in the interview was to read a comic book. Third, all boys who failed to answer one or more questions in Part II of the questionnaire were left out, since these questions were to be repeated after the comic book reading. This procedure left 144 boys (H-H: 29, H-L: 38, L-H: 36, L-L: 41) from whom the final selection was made. The final groups were chosen in such a way as to match as closely as possible the distributions on the following characteristics: religion, father's occupation, education of parents, IQ, and school achievement as measured by average grade for the year 1954–55. Of the 100 boys specified in this way, two could not be interviewed: one because he had moved and had entered one of the two schools to which we did not have access; the other because his parents had been disturbed by the questionnaire and the school authorities requested the removal of his name from the list. Two other boys were substituted with social characteristics as close to the original two as possible, and from the same exposure-problem groups. Each selected boy was interviewed individually during the month of May, 1955, six to ten weeks after he had filled out the questionnaire.

APPENDIX B

QUESTIONNAIRE AND MARGINALS BY SEX

The questionnaire consists of four parts: face sheets—social characteristics; Part I—psychological and some cognitive characteristics; Part II—cognitive characteristics; Part III—exposure habits. This appendix presents Parts I and II of the questionnaire. The other parts may be found in the dissertation copy (Bailyn, 1956, App. A).

The figures next to the response categories represent the percentage of children choosing that category. The figure on the left represents the percentage of boys giving that response; the one on the right, the percentage of girls. Percentages do not always add up to 100% because some children did not answer certain questions or circled more than one response.

In questions asking for more than one choice, a number of pairs of figures are given: the pair at the left represents the percentage of children choosing that category as their first choice (boys to the left of girls), the second pair represents the second choice, the third pair the third choice, etc.

Classification of answers to open-ended questions are given immediately following the questions. Response categories to these questions are neither numbered nor identified by letters. Again, the percentage of boys' answers in each classification appears to the left of that of girls' answers.

PART I

1. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

66,87	1. Go on an outing or picnic for the afternoon with your friends.
33,13	2. See a movie with your friends.

2. Suppose that, just by wishing, you could be any age at all. Put a circle around the number in front of the age you would most like to be:

6,1	1. Under 5
3,4	2. 5 to 9
29,28	3. Over 9 but under 13
36,46	4. 13 to 19
25,21	5. 20 or over

3. In this question put a "1" in front of your first choice, a "2" in front of your second choice, and a "3" in front of your third choice.

If you were going to the circus would you want to go—

52,25	23,43	10,15	A—with your father?
12,15	16,10	21,19	B—with your best friend?
4,6	9,3	30,37	C—with your brothers or sisters?
9,10	9,9	14,12	D—with a group of friends?
18,41	39,33	15,11	E—with your mother?
4,1	4,1	7,3	F—by yourself?

4. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

42,44	1. Be asked to join a club.
55,54	2. Start a new club.

5. Circle the number of the answer that fits you best:

5,10	1. I am too tall.
18,10	2. I am too short.
76,80	3. I am just about right.

6. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

48,68	1. Read a book.
51,32	2. Read a comic book.

7. What would you most like to be when you are grown up? Put down the 3 types of work you would most like to do:

2,3	2,2	1,2	occupations involving animals

8,1	8,2	10,1	"cops and robbers"
7,4	7,5	7,3	air, sea, and other adventure occupations
7,2	6,3	6,1	armed services
3,1	4,1	5,0	outdoor occupations
11,0	11,0	8,1	athletics
1,10	1,10	2,10	entertainment
1,2	1,2	1,3	arts
11,0	8,0	5,0	builder-engineer
6,1	6,0	7,2	scientist
16,3	8,1	8,5	top professional
4,0	4,0	5,1	business and managerial
1,2	1,0	1,0	religious occupations
0,18	1,10	2,15	teacher
3,24	2,18	2,12	lesser professional
1,0	2,0	3,0	salesman
2,2	3,3	2,3	service
1,17	1,20	1,16	clerical and sales
11,0	16,0	11,0	worker, all skills
0,9	0,11	0,14	housewife
2,1	3,3	5,3	other

8. Circle the number of the answer that fits you best:

11,19 1. I am too fat.
12,10 2. I am too thin.
77,70 3. I am just about right.

9. How many friends do you have? Put a circle around the number of the answer that fits you best:

6,2 1. I don't have very many friends and like it that way.
8,9 2. I don't have very many friends and wish I had more.
9,8 3. I have a few good friends and don't want any more or less.
11,16 4. I have a few good friends and would like to have more.
41,38 5. I have many friends and enjoy it.
3,4 6. I have many friends but would rather have just a few good ones.
19,23 7. I have many friends and would like to have even more.

10. Suppose you got a bad mark on your report card. How would you feel? Circle the number in front of the answer that fits you best:

95,98 1. I would feel it was my fault because I didn't work hard enough.
3,1 2. I would feel that mostly it was not my fault.

11. How often do you go to the homes of other kids in your class? Put a circle around the number of the answer that fits you best:

36,27 1. Very often (almost every day).
34,42 2. Sometimes (once a week or so).
30,30 3. Hardly ever.

12. A boys sits down at the table to eat his soup. His mother says, "It's too bad that the soup is cold." What do you think the boy answers? Circle the number of the answer you think the boy gives to his mother.

The boy says:

12,9 1. "Can't you heat it some?"
39,31 2. "That's all right."
48,60 3. "I should have started eating sooner."

13. Do you ever get spanked nowadays? Put a circle around the number of the answer that fits you best:

4,1 1. Yes, quite often.
33,23 2. Sometimes.
62,76 3. No, hardly ever.

How old were you the last time you were spanked?

4,7 4 years or less
5,5 5
6,7 6

6,10 7
7,14 8
15,16 9
24,20 10
20,11 11 years or over

14. Suppose you and your friend were talking to each other and then you find out that your friend is unhappy about something that was said. Would you feel this was your fault? Circle the number in front of the answer that fits you best:

23,15 1. I would feel it was definitely my fault.
64,78 2. I would feel it might be my fault.
7,6 3. I would feel it probably was not my fault.
5,1 4. I would feel it definitely was not my fault.

15. How much difference does it make to your parents what marks you get in school? Circle the number in front of the answer you think is right:

93,80 1. It is very important to them for me to get good marks.
7,19 2. They are pretty much satisfied with any marks I get.

16. How often do other kids in your class come to your house? Put a circle around the number of the answer that fits you best:

31,21 1. Very often (almost every day).
37,46 2. Sometimes (once a week or so).
31,33 3. Hardly ever.

17. Do your parents ever keep you from doing the things you want to do? Put a circle around the number of the answer that fits you best:

11,5 1. Yes, quite often.
74,76 2. Sometimes they do.
15,19 3. They hardly ever do.

18. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

30,17 1. Go to someone else's party.
68,83 2. Give a party.

19. Circle the number in front of the person who punishes you when you do something wrong:

39,63 1. Usually my mother.
55,35 2. Usually my father.

Does anyone else punish you? If yes, write here who it is:

9,11 Other parent
19,13 Other
71,75 No one else

20. A boy is playing marbles with his friend. The friend says, "I'm sorry I pushed your marble by mistake." What do you think the boy answers? Circle the number of the answer you think the boy gives to his friend.

The boy says:

45,53 1. "I should have moved it out of your way."
43,40 2. "Let's keep playing."
11,6 3. "Can't you be more careful?"

21. Do you think your parents try to find out how you feel about something if you and they disagree? Circle the number in front of the answer you think is right:

43,46 1. Yes, almost always.
41,42 2. Sometimes they do.
16,11 3. Usually they don't.

22. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

49,53 1. Play a game indoors.
49,46 2. Watch television.

23. Suppose you could just have 3 of the wishes below, which would you want to come true? Read them all, then put a "1" in front of your biggest wish, a "2" in front of your second biggest wish, and a "3" in front of your third wish.

I would like:

4,9	7,8	5,10	A–To have more friends.
17,4	15,3	11,2	B–To be stronger than I am now.
20,32	13,16	10,14	C–To see the world.
5,5	3,3	3,1	D–To get along better with my father and mother.
4,10	3,8	4,6	E–To be thinner than I am now.
6,3	9,4	10,5	F–To go to a ball game.
3,7	7,14	7,12	G–To be better looking than I am now.
1,0	1,2	2,0	H–To have my father and mother be more like other people's parents.
0,1	5,4	6,2	I–To be taller or shorter than I am now.
2,4	4,7	4,12	J–To play a musical instrument.
1,2	1,4	4,6	K–To get along better with other kids.
16,2	18,6	15,5	L–To have a private plane.
1,1	1,3	2,1	M–To have my father and mother pay as much attention to me as they do to my brothers and sisters.
20,19	13,17	17,21	N–To be smarter than I am now.

24. Which of these two things would you usually rather do? Circle the number of the one you would rather do:

20,14	1. Decide yourself which games to play.
66,80	2. Let others decide which games to play.

25. Suppose you are talking to a friend and something the friend says makes you unhappy. Would you feel that this was your friend's fault? Circle the number in front of the answer that fits you best:

14,4	1. I would feel it was definitely my friend's fault.
46,52	2. I would feel it might be my friend's fault.
29,35	3. I would feel it probably was not my friend's fault.
11,8	4. I would feel it definitely was not my friend's fault.

26. Suppose that just by making a wish you could change yourself into any of the people below. Read them all, then put a "1" in front of the person you would most want to be changed into, a "2" in front of your second choice, and a "3" in front of your third choice.

I would like to be changed into a:

3,0	2,0	4,0	A–king
6,5	6,4	5,2	B–cowboy
10,45	8,12	8,15	C–movie star
9,1	5,3	6,1	D–president
0,1	2,1	0,1	E–prince
10,1	12,7	9,4	F–detective
4,5	6,13	6,15	G–reporter
7,1	9,2	8,5	H–inventor
14,1	6,3	6,4	I–Superman
2,0	2,1	4,2	J–space cadet
0,5	0,15	0,14	K–queen
12,13	6,6	5,6	L–animal breeder
8,2	12,6	13,4	M–explorer
0,15	0,21	0,17	N–princess
11,2	18,3	17,5	O–FBI agent
3,0	5,0	4,0	P–knight
1,0	0,1	2,1	Q–person from Mars

PART II

1. Do you think it is important to know if a person is good or bad before having anything to do with him?

86,86	1. I think it is very important.
14,14	2. I don't think it is very important.

2. Why do some people become bums?

52,44 1. Because they are naturally lazy.
47,55 2. For some other reason.

3. Do you think this country is in danger?

18,11 1. I think it is in great danger.
41,41 2. I think it is in a little danger.
23,31 3. I don't think it is in great danger.
17,16 4. I don't think it is in any danger.

4. It is possible to tell by a person's face whether he is honest or dishonest.

36,38 1. I think this is true.
62,61 2. I think this is false.

5. How often does something stop a person from doing the things he wants to do?

5,4 1. Almost always.
38,35 2. Quite often.
45,51 3. Not usually.
10,9 4. Hardly ever.

6. Circle the number of the person you think gets along better in this world:

94,95 1. A weak but nice person.
5,5 2. A strong but not quite so nice a person.

7. Why do people who commit crimes do so?

22,20 1. Because they are naturally bad.
77,79 2. For some other reason.

8. Do you think there is danger that people on this earth will be harmed by beings from another planet?

4,1 1. Yes, quite likely there is.
45,39 2. It is always a possibility.
32,48 3. Most likely not.
18,12 4. No, definitely not.

9. Lazy people are all alike and quite different from people who work hard.

74,70 1. I think this is true.
25,30 2. I think this is false.

10. Can most people be trusted?

21,21 1. Almost all people can.
55,54 2. Quite a lot of people can.
22,24 3. Only some people can.
1,1 4. Hardly any people can.

11. Do criminals ever become good?

78,82 1. Sometimes they do.
20,17 2. No, they usually don't.

12. Some people always get pushed around. Why is this?

26,14 1. Because they are naturally weak.
74,86 2. For some other reason.

13. Do people live here who are really enemies of this country?

18,7 1. Yes, many do.
44,35 2. Yes, some do.
21,34 3. No, not many do.
17,23 4. No, hardly any do.

14. People are either all good or all bad.

24,25 1. I think this is true.
75,74 2. I think this is false.

15. Is it hard for a person to be what he wants to be in life?

20,10 1. It is very hard.
40,44 2. It is quite hard.
32,40 3. It isn't very hard.
8,5 4. It isn't hard at all.

16. Are good people usually strong?

6,5 1. Yes, usually.
93,95 2. Not necessarily.

17. Why do some people tell lies?

14,13 1. Because they are naturally dishonest.
85,86 2. For some other reason.

18. Do you think something may destroy the world some day?

25,12 1. Yes, I definitely think so.
43,51 2. I think it might happen.
22,32 3. Most likely not.
10,5 4. No, I definitely don't think so.

19. There are only two types of people in the world, the weak and the strong.

19,19 1. I think this is true.
81,81 2. I think this is false.

20. Can a person usually find help if he needs it?

39,40 1. He almost always can.
53,55 2. Quite often he can.
6,4 3. He seldom can.
0,0 4. He hardly ever can.

APPENDIX C

SCORES AND CODING PROCEDURE

Scores

In the derivation of these scores, when a relevant question was left blank by any child a score was assigned to him on this question on the basis of his responses to the other items involved, as well as on the general distribution of answers to the question.

Exposure score. The exposure score was developed on the basis of the means and standard deviations of the frequency distributions for television, movies, and comic books of a sample of 106 boys. These data are presented in Table 28.

The frequencies given in the questionnaire were first changed into deviation scores. To make them more manageable, these scores were then subjected to the transformation 5x + 15.5. The resulting values are given in Table 29.

The exposure score consists of the sum of the appropriate values for television, movies, and comic books.

Problem score. (Part I, Questions 2, 3, 5, 8, 9, 11, 15, 16, 17, 21, 23. Questions 2, 3, 9, and 23 are modified versions of items given by Rogers [1931, pp. 10, 16–18].) Three areas of problems were used to arrive at this score: self, peer, family. The scoring procedure is presented for each area separately.

TABLE 28

MEANS AND STANDARD DEVIATIONS OF FREQUENCY DISTRIBUTIONS TO PICTORIAL MEDIA OF SAMPLE FROM WHICH SCORE WAS DERIVED

	Television	Movies	Comic Books
Mean	2.28	4.15	3.51
Complement of mean to Base 7	4.72	2.85	3.49
Standard deviation	1.20	1.25	1.65

TABLE 29

VALUES USED IN EXPOSURE SCORE

Response No.	Television	Movies	Comic Books
1	21	28	23
2	17	24	20
3	13	20	17
4	8	16	14
5	4	12	11
6	0	8	8

SELF AREA (Nos. 2, 5, 8, 23)

Response	Score
Item No. 2	
1	4
2	3
3	0*
4	1**
5	2

* Unless age outside this range, then count as 1.

** Unless age in this range, then count as 0.

Response	Score
Item No. 5	
1	Boys—3; Girls—5
2	Boys—5; Girls—3
Item No. 8	
1	5
2	3

Item No. 23

	1st Choice	2nd Choice	3rd Choice
B	5	3	1
E	5	3	1
G	5	3	1
I	5	3	1
N	5	3	1

PEER AREA (Nos. 3, 9, 11, 16, 23)

Response	Score
Item No. 3	
Chooses F but not B or D	3
Item No. 9	
1	3
2	5
4	3
6	1
7	1
Item No. 11	
3	5
Item No. 16	
3	5
Item Nos. 11 and 16 Combined	
3 and 1 respectively	2
3 and 2 respectively	1

Item No. 23

	1st Choice	2nd Choice	3rd Choice
A	5	3	1
K	5	3	1

FAMILY AREA (Nos. 3, 11, 15, 16, 17, 21, 23)

Response	Score
Item No. 3	
Chooses E but not A (if father at home)	Boys—3; Girls—0
Chooses A but not E (if mother at home)	Boys—0; Girls—3
Chooses A, E, F (if child has older siblings)	3
Chooses B, D, F	2
Item Nos. 11 and 16 Combined	
1 and 3 respectively	2
2 and 3 respectively	1
Item No. 15	
1	3
Item No. 17	
1	5
Item No. 21	
2	2
3	5

Item No. 23

	1st Choice	2nd Choice	3rd Choice
D	5	3	1
H	5	3	1
M	5	3	1

The score for each area consists of the sum of the scores for all the relevant questions. The sum of these three separate scores comprises the total problem score.

Direction of punitiveness. (Part I, Questions 10, 12, 14, 20, 25. Questions 12 and 20 are based on Rosenzweig's picture-frustration test for children [Rosenzweig, Fleming, & Rosenzweig, 1948, pp. 167–170].)

Response	Score
Item No. 10	
2	4
Item No. 12	
1	4
2	2
Item No. 14	
2	1
3	2
4	3
Item No. 20	
2	2
3	4
Item No. 25	
1	4
2	3
3	2
4	1

Coding Procedure

The reliability of coding was determined by having a different person from the one who did the coding recode 51 questionnaires. These questionnaires were chosen so that each class participating in the study had at least two representatives—one boy and one girl. Within a class, the questionnaires were picked at random. The percentage of disagreements between the original coding and the recoding of these questionnaires gives the percentage of coding error; 100% minus this percentage indicates the reliability.

Occupation of father. The coding procedure for father's occupation was based on an occupational code developed at the Laboratory of Social Relations, Harvard University, for a study on the aspirations of high school boys, under the joint direction of Florence Kluckhohn, Talcott Parsons, and Samuel Stouffer. The categories follow:

1. Top professional; major proprietary or managerial.
2. Lesser professional; average proprietary or managerial; major sales and clerical.
3. Minor proprietary or managerial; minor sales and clerical.

An occupation that was clearly of a white-collar character but where insufficient information was given to code into any of these three categories was classified as

4. White collar—unascertainable.

Working occupations were classified, by means of the *Dictionary of Occupational Titles,* into the following categories:

5. Skilled
6. Semi-skilled
7. Unskilled

Service positions such as policeman, hairdresser, bus driver were classified into the skilled category; such occupations as barbers, milk men, waiters were classified with the semi-skilled workers. Occupations clearly of a worker or service character that were not classifiable into one of these three categories were coded as

8. Worker—unascertainable.

Occupations that were so insufficiently described that they could be either white collar (Categories

1–4) or worker (Categories 5–8) were coded
9. Unascertainable.

Reliability of coding: When any deviation in these nine occupational categories is counted, the coding error was 24%, the reliability 76%. When only three classifications are used—white collar, worker, unascertainable—the reliability of coding was 92%.

Occupational choice, realism. Occupational choices were classified as realistic or unrealistic on the basis of the socioeconomic status of the family and the capabilities of the child. The coding was done on the basis of an estimate of the probability of a child in a given situation and with certain capabilities actually attaining the desired occupational status. Many individual decisions had to be made; the following general rules guided the coding.

Realistic:

- Top professional—if father in top two white-collar categories and IQ of child ≥ 106 (in top two thirds of the IQ distribution) and grades ≤ 2.35 (in top two thirds of distribution); if father not in top two white-collar categories and IQ ≥ 122 (top third) and grades ≤ 1.45 (top third).
- Lesser professional—if IQ ≥ 106.
- Business—unless clearly out of range of family status.
- Worker—unless father in top two white-collar categories.
- Service—unless father in top two white-collar categories.
- Religious choices—unless child does not go to Sunday School.

Unrealistic:

- Air and sea occupations, animal, art, athletics, entertainment, outdoor occupations—unless parents engaged in these activities.
- Policeman and fireman—if father in top two white-collar categories.

Some choices were unascertainable in this respect and were not classified into either of these categories.

A first occupational choice that was realistic was given a score of 3, a second 2, and a third 1. Hence, the realism scores range from 0 (all unrealistic) to 6 (all realistic) for children who gave three choices that were ascertainable in this respect.

Reliability of coding: When any difference in realism score is counted as an error, there was a 16% error of coding, 84% reliability. When only the first choice and the categories of realistic, unrealistic, or unascertainable are used, the coding reliability was 90%.

Occupational choice, active nature of realistic choices. All realistic choices were classified as active or not on the basis of a comparison with the father's occupation for boys and with the mother's for girls. We present the types of combinations that were considered active occupational choices. On the basis of these rules the individual decisions were made.

Active occupational choices for boys:

Father	*Son*
nonprofessional	professional of any kind
non-top professional	top professional
managerial	proprietal
top professional	top business
second white-collar category	first white-collar category
lowest white-collar category	top two white-collar categories
worker	white collar
unskilled or semi-skilled worker	skilled worker or service occupations

Active occupational choices for girls:

All the above rules apply when "mother" is substituted for "father," also the following:

Mother	*Daughter*
untrained service	trained service
housewife	anything but housewife

Some realistic choices were unascertainable in this respect.

Each child was given a score on activeness on the basis of the percentage of his ascertainable realistic choices that were active. Children with no realistic choices that were ascertainable in this respect were given no score on activeness.

Reliability of coding: When all differences in percentages are counted as errors, the coding error was 18%, reliability 82%. Use of the categories of all, some, or no active choices gives a reliability of 94%.

Occupational choice, media orientation of unrealistic choices. Unrealistic choices were classified as media oriented or not so oriented, on the basis of the estimated frequency with which acquaintance with such occupations stems from material in the mass media as opposed to other sources. The following rules indicate the type of decisions made.

Media Oriented	*Non-media Oriented*
movie star	professional
artist on a medium	musician, author
cowboy or cowgirl	most outdoor occupations
detective, FBI agent	
space cadet or ranger	

Animal and athletic choices were considered unascertainable for this purpose.

Each child was given a media-orientation score on the basis of the percentage of his ascertainable unrealistic choices that were media oriented. Children with no unrealistic choices that were ascertainable in this respect were given no score on media orientation.

Reliability of coding: When all differences in percentages are counted as errors, the coding error was 16%, reliability 84%. Use of the categories all, some, or no media-oriented choices gives a reliability of 92%.

Bureau Ednl. Psy. Resea
DAVID HA AINING COLLEGE
Dated

Vol. 73, No. 2 Whole No. 472, 1959
Psychological Monographs: General and Applied

THE EFFECTS OF EARLY EXPERIENCE ON SUBSEQUENT EMOTIONALITY AND RESISTANCE TO STRESS[1]

ROBERT ADER[2]
Cornell University

Although there is a steadily increasing number of studies dealing with the effects of early experience on behavior, relatively few of these investigations have been concerned with emotionality. The research reported here is concerned with the effects of various types of early experiences on emotionality and resistance to stress in the rat. It was prompted by a reinterpretation of previous findings in this area, hence the following review is particularly pertinent.

One should mention at the outset that despite the amount of literature purporting to deal with the effects of early experience, there are, in fact, very few studies which actually bear on this subject. Unless one manipulates both young and old individuals in the same manner, one is not investigating the effects of *early* experience but only the effects of *previous* experience. Unless it is known that a particular type of experience does not effect the same changes in the subsequent behavior of mature individuals that it does in the subsequent behavior of individuals subjected to this experience early in life, or unless the differences between experimental and control subjects manipulated early in life are greater than those between experimental and control subjects manipulated at maturity, one can only conclude that prior experience exerts an influence on behavior. A majority of the studies cited below (as well as others in this area) have failed to use or mention this distinction.[3]

Early Experience and Emotionality

One of the earliest studies designed specifically to investigate the effects of "early" experience on subsequent emotional behavior was that of Hall and Whiteman (1951). Reacting against J. E. Anderson's (1948) statements contradicting the psychoanalytic emphasis on early trauma, these experimenters tested the hypothesis that "*subjecting the infant organism to intense stimulation will result in emotional instability in later life*" (Hall & Whiteman, 1951, p. 61). Beginning at 4 days of age, mice were subjected to four 2-min. periods of a loud, high-frequency bell. At 30 to 40 days of age, these mice and unstimulated controls were tested in an open-field situation. At 70 to 80 days of age, Stone's (1932) stovepipe test of emotionality was employed, and at 100 to 110 days of age the open-field test was repeated. On the basis of defecation, urination, and latency scores, found to be indicative of emotionality (E. E. Anderson, 1938; Hall, 1934; Hunt & Otis, 1953; Tryon, Tryon, & Kuznets, 1941b), it was concluded that animals subjected to intense stimulation during infancy are emotionally unstable in later life.

Although the differences on the first open-field test were significant, the authors themselves point out that on the stovepipe test only six animals could be described as having been influenced by the early experience. The remaining 15 experimental mice behaved normally. On the final open-field test the observed differences were not statistically significant, so that the results as a whole remain equivocal.

Griffiths and Stringer (1952) exposed groups of rats to a variety of noxious stimuli from birth to 21 days of age. A control group receiving "approximately the same treatment" (custodial care?) was not exposed to any of these stimuli. Beginning at 60 days of age, all groups were observed in learning situations and in an open-field situation. No differences in learning ability or in emotionality were observed between any of the groups.

---

[1] This research represents a portion of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Cornell University, 1957. Thanks are due H. S. Liddell and G. W. Boguslavsky under whose direction this study was conducted.

[2] Present address: Department of Psychiatry, University of Rochester School of Medicine and Dentistry, Strong Memorial Hospital, Rochester, N. Y.

[3] For the reader interested in reviewing this literature it should be mentioned that studies published subsequent to the writing of this paper have taken into account these considerations.

Infant mice were also exposed to noxious stimulation (electric shock) by Stanley and Monkman (1956). Here again, subsequent testing in an open field revealed no differences in the amount of defecation—or activity[4]—between animals receiving escapable shock, those receiving unescapable shock, and those placed into the apparatus without receiving shock.

Weininger (1956) has investigated the effects of "early" handling on subsequent emotional behavior. Animals handled for 10 min. each day for 21 days following weaning were found to be more active in an open-field situation than controls. The open field employed was 6 ft. in diameter and measured off into four concentric circles. "Gentled" animals also showed a greater incidence of entries into the innermost circles; nongentled animals tending to move about the outer walls of the apparatus (thigmotaxis), which is the usual behavior in this species.

Scott (1955) found no differences in the activity (in a Miller-Mowrer box) between animals which were handled, exposed to electric shock, or ignored for the 21 days following weaning. In this case the manipulation lasted only 3 min. per day. The discrepancy between Scott's results and Weininger's is probably not due to the difference in the amount of manipulation, however, since it has been demonstrated that animals handled for 3 min. per day for 30 days following weaning are less emotional than controls as measured by activity in an open field (Ader, 1957). Before testing in the open field, these animals were observed in an "emergence-from-the-cage" test,[5] the time required to emerge from the open home cage being indicative of emotionality or "timidity" (E. E. Anderson, 1938; Bindra, 1948; Hunt & Otis, 1955). During four such tests, only two of nine handled animals failed to leave their cage within a 15-min. period. Only one of eight nonhandled animals emerged from its cage, and this was on the last of the four tests. Again, handled animals prove to be the less emotional.

Results comparable to those described above were obtained by Hunt and Otis (1955). Although there were no differences in the amount of defecation in an open field, animals handled from the seventh through the twenty-first day of life displayed less emotional behavior, as measured by the number of animals emerging from a standard starting cage, than did controls given a restricted environment.

More recently Doolittle and Meade (1957) investigated the effects of handling rats at different periods of time. One group was given 10 min. of handling per day for three weeks following weaning while a second group was given the same manipulation beginning at 45 days of age. The testing of all groups began at 71 days of age. Whereas there was no evidence for the efficacy of early versus late handling, both handled groups, as measured by the amount of defecation which occurred during periodic weighing, were somewhat less emotional than controls. After depriving the animals, activity was recorded in a T maze in which they had previously received a 2-min. adaptation period. Nonmanipulated control animals showed the greatest amount of activity. This result is at odds with the results of previously cited studies (Ader, 1957; Weininger, 1956). A possible explanation lies in the fact that, as Doolittle reports, the handled groups were slightly heavier than the controls and consumed more food. This would lead one to suspect differential effects from the deprivation schedule which was imposed.

Of particular relevance to the present thesis are the discrepant results cited above. While there is general agreement that handling decreases emotional reactivity (Ader, 1957; Hunt & Otis, 1955; Weininger, 1956), there is not the same agreement between studies of the effects of noxious stimulation. Also, except in a single instance (Doolittle & Meade, 1957), the lack of appropriate control groups of mature animals prohibits any definitive statement concerning the effects of *early* experience on emotionality.

Early Experience and Avoidance Learning

Avoidance learning as a function of "early" experience has also been investigated, and in each instance the early experience has been electric shock. In the study by Stanley and Monkman (1956) one group of mice was exposed to shock that could be terminated by movement of the animal. A second group received shock that could neither be avoided nor escaped, and a third, placed into the apparatus, did not receive shock. As previously mentioned, no differences in emotionality were observed. In later avoidance training, with electric shock as the unconditioned stimulus, the group having received early experience with escapable shock was superior. The authors describe these results in terms of a positive transfer effect.

Chevalier and Levine (1955), and Levine, Chevalier, and Korchin (1956) also gave animals "early" experience with inescapable electric shock. In this study one group of rats was given daily shock from birth to 21 days of age. A second group was placed on the grid but did not receive shock, and a third group was completely ignored during this period. When given avoidance training at 60 days of age, the unmanipulated controls were inferior to the other two groups on all measures. There were no differences between the shocked and manipulated groups in the number of trials to criterion or the number of initial errors (trials to the first avoidance response). The total number of errors, however, was greater for shocked animals than for the no-shock, handled group. One hypothesis suggested by these authors to account for this latter result is that shock in infancy may interfere directly with the cognitive functioning necessary for adequate performance in this situation. Such an interpretation, however, is at odds with the remainder of their data, since shocked and manipulated groups in the number of mals on the other measures and shocked animals are still superior to unmanipulated controls. An alternative hypothesis is that "early trauma leads to a heightened anxiety drive which persists into adulthood" (Levine et al., 1956, p. 490), and that the poorer performance of shocked animals is thus an effect of intense drive. Again, these animals show *superior* performance in contrast to unmanipulated controls, and the Stanley and Monkman results are evidence that such treatment can have a positive effect on subsequent performance.

---

[4] W. C. Stanley. Personal communication, Nov. 21, 1956.

[5] Described by the author in unpublished data.

The hypotheses of Chevalier, Levine, and Korchin are further contradicted by the results of the studies described below. In one (Steckle & O'Kelly, 1941), beginning at 20 days of age rats were given 35 days of experience with electric shock in that half of their cage where water was to be found. When tested later (at 70 days) in a situation where they had to cross an electrified grid to obtain water, the shocked group had a greater mean number of grid crossings as well as a greater percentage of animals that crossed the grid to water than did an untreated control group. According to the authors, " . . . the *early* experience with electric shock has produced an adaptation or a superior learned means of adjustment to the shock barrier situation" (p. 8, italics supplied), which has been carried over to the test situation.

It seems pertinent at this point to reiterate the fact that, unless one manipulates both young and old individuals in the same manner, one is not investigating *early* experiences but only the effects of *previous* experiences. That an individual's past experiences influence his present behaviors is well documented. That *early* experience is particularly influential remains to be conclusively demonstrated.

One further study of the effects of early experience on avoidance conditioning is especially worthy of mention because it is one of the few studies controlling for *early* experience. Baron, Brookshire, and Littman (1956) subjected rats to a strong electric shock for 6 min. on two consecutive days at either 20 days, 36 days, or at maturity. A control group was not shocked. Animals were later trained in either an escape or avoidance situation under either high or low shock as the unconditioned stimulus. In accordance with Doolittle and Meade's (1957) results on handling, and in contrast to the results of Hunt (1941) on hoarding behavior, Hymovitch (1952) and Hebb (1947) on problem solving, and Wolf (1943) on the inhibition of function in neurosis, all of whom found *early* experience to be crucial, these authors report that the animals subjected to the electric shock *at any time* prior to the learning situation were superior to controls in escape learning under both high and low shock and in avoidance learning under high shock. Although not significant, the results of avoidance learning under low shock were in the same direction. Traumatization, in infancy only, resulted in slower extinction of the escape learning under high shock. "The results are interpreted as: (1) substantiating the hypothesis that traumatization will affect subsequent learning *in which the traumatizing agent recurs;* (2) substantiating the hypothesis that effects of trauma are long lasting; (3) not strongly substantiating the hypothesis that traumatization at an early age has broader consequences than traumatization at maturity."[6]

Early Experience and Physiological Responses

In studies designed specifically to determine the effects of "early" experience on growth characteristics, Weininger and his associates (McClelland, 1956; Weininger, 1954, 1956; Weininger, McClelland, & Arima, 1954) observed that handled animals weighed significantly more than unmanipulated controls. With but one exception (Scott, 1955), similar findings have been reported by other investigators (Bernstein, 1952; Ruegamer, Bernstein, & Benjamin, 1954).

Griffiths and Stringer (1952) report normal growth curves for groups of animals exposed to a variety of noxious stimuli during early life. Such groups did not differ from control animals. In contrast to these results, two of the groups manipulated by Herrington and Nelbach (1942) have relevance here. These authors found that by 210 days of age a group of "disturbed" animals (those exposed to "intermittent bell ringing, air blasts, and cage vibration once each hour") showed a greater body weight than controls (those maintained in a room with a temperatue of 83° F.). It is not known, however, at what age this treatment was begun nor for how long it lasted.

Weininger (1953) has also found that the mortality rate in the rat is affected by "early" experience. Adult animals were subjected first to 60 hr. without food or water which was then continued until 120 hr. had elapsed. Under this stress all nongentled animals eventually succumbed. Those animals handled for 10 min. per day during the 21 days following weaning survived. These results confirm the earlier findings of Hammett (1921, 1922) who observed a decreased mortality rate in previously tamed animals subjected to an operation for removal of the parathyroid gland.

Again to the contrary, Scott (1955) found no differences in the adult mortality rate of handled animals, those exposed to electric shock, and animals ignored throughout infancy. The stress situation imposed was the injection of 100 mg./100 gm. body weight of thiourea. According to Bernstein (1956), such a dosage would be extreme and one may infer that no manipulation could prevent death from such a dosage. Furthermore, Scott administered these injections to mature animals (136 days). Bernstein (1956), referring to a study as yet unpublished, obtained contradictory results with young animals, and it has been reported elsewhere (Mackenzie & Mackenzie, 1943) that the susceptibility of the rat to the toxic effects of thiourea increases with age.

---

[6] A. Baron. Personal communication, Oct. 8, 1956 (italics supplied).

Of particular relevance to the present thesis is the work of Noble (1943) and Noble and Collip (1942). These investigators have devised a method for producing graded amounts of trauma in small animals. "The animal, after taping of the paws, is placed in a circular drum which has two projections on the side. During rotation of the drum the animal is alternately carried a short distance up the side on the projection and then dropped. The amount of trauma, therefore, is directly related to the number of turns of the drum, and is reflected on the mortality curve" (Noble, 1943, p. 346). That is, the greater the number of turns the greater the trauma, and the greater the trauma the sooner animals die. Subjecting animals to 1000 turns results in mortality in 98% of the cases tested. Characteristically, death ensues following termination of the trauma rather than during the tumbling. Repeated experimentation has shown that rats may be made resistant to this trauma by subjecting them to increasing amounts of sublethal tumbling in the drum. Within a period of 12–14 days, an animal may be raised from 200 to 1000 turns and survive. Being subjected to 1000 turns every week thereafter still fails to cause death. It has not yet been determined if animals made resistant to trauma in the Noble-Collip drum maintain this resistance when exposed to other forms of stress.

Griswold and Gray (1956), having observed that the administration of various autonomic drugs prior to stress serves to reduce mortality, conclude that a diminution of activity in part or all of the autonomic nervous system serves to protect an animal from stress. Tumbling in the Noble-Collip drum elicits the classical emergency pattern of responses and, "It is conceivable that such repetition of sympathetic discharge could result in a gradual alteration in the character of the process so that on subsequent exposure to an ordinarily lethal challenge, the sympathetic system reacts with less violence, thereby causing the survival of the animal. Alternatively, it is possible that the effector organs activated by the autonomic system may become less responsive to a given strength of discharge after having been repeatedly stimulated" (Griswold & Gray, 1956, p. 1).

Based on this reasoning, Griswold and Gray chose electroconvulsive shock (ECS) as the means by which autonomic stimulation could be achieved. Following a series of ECS treatments, rats were subjected to tumbling trauma in the Noble-Collip drum. It was found that one ECS for 4 days or four ECS in a space of 1 or 2 days resulted in an increased resistance to the subsequent trauma. Ten and 20 ECS's produced only a small additional increase in resistance.

After observing his groups in the behavioral situations previously discussed, Weininger (1956) and Weininger et al. (1954) subjected handled and nonhandled animals to a severe stress situation. Animals were immobilized in gauze and adhesive tape and remained so for 48 hr. without food or water. At the end of this period they were sacrificed and autopsies were carried out. Nongentled animals showed greater damage and distension of blood vessels of the heart, and suffered more bleeding points throughout the gastrointestinal system. There were no differences in the weights of the liver, kidneys, or pancreas, but the adrenals were heavier in the nongentled animals. The organ weights of nonstressed handled and nonhandled animals did not differ.

In attempting to account for these results, Bovard (1954) has suggested that a reduction in sympathetic activity and decreased ACTH output from the pituitary, effected by a change in hypothalamic functioning, accounts for the hypo-activity of the adrenal glands of gentled animals. That an increase in resistance is effected by a decrease in autonomic activity was also hypothesized by Griswold and Gray (1956).

While similar to Weininger's handled animals in the direction of the body weight difference, Herrington and Nelbach's (1942) "disturbed" animals differed from "gentled" animals in that autopsy showed the adrenal glands to be *heavier* than those of the control group.

The relationship between the adrenal glands and reactions to stress is well documented (Selye, 1950), adrenal size being indicative of the gland's capacity for sustained secretion (Hartman & Brownell, 1949). On a behavioral level, adrenal size has been found to be positively correlated with emotionality (Hall, 1939; Yeakel & Rhoades, 1941), as measured by the open-field test.

The Hypothesis

The present thesis derives from the writer's interpretation of some of the research findings discussed above. The present section, then, is concerned with an attempt to organize some seemingly discrepant results with regard to the effects of early experience on subsequent emotional behavior.

Concerning the effects of specific types of "early" experience on subsequent emotionality in situations which are not specific to the early experience, we have seen that there is general agreement that handling decreases emotional reactivity (Ader, 1957; Hunt & Otis, 1955; Weininger, 1954). There is not the same agreement between studies of the effects of noxious stimulation during early life. It has been variously reported that "emotionalized" animals are emotionally unstable (Hall & Whiteman, 1951), do not differ from controls (Griffiths & Stringer, 1952; Stanley & Monkman, 1956), and are less emotional than control animals (Ader, 1957).

The research by Weininger and his associates (McClelland, 1956; Weininger, 1954, 1956; Weininger et al., 1954) has shown that handling, as a variable experienced in early life, is capable of altering the rate of growth, structure and behavior of the rat. Animals thus treated show less emotional behavior and appear more capable of withstanding severe stress. These findings have been supported both physiologically (Hammett, 1921, 1922; Ruegamer et al., 1954; Ruegamer & Silverman, 1956) and behaviorally (Ader, 1957; Bernstein, 1952; Karn & Porter, 1946).

The reference to these animals as "gentled" (Weininger, 1956) connotes a personality trait, which, although operationally defined, suggests itself as a correlate of the physical and behavioral characteristics observed. While such a relationship may be the case, such a conclusion would be premature at present, and recent evidence on the effects of other types of previous experience on emotionality (Ader, 1957) suggests that handling is not the only variable that can effect a heightened reaction threshold to emotion-provoking situations. In this study, one group of animals was handled for 3 min. per day for 30 days following weaning. A second group was subjected to an emotion-provoking situation equal in distribution and duration to the handling procedure. A control group was completely ignored during this period. As measured by an open-field test, the behavior of mature animals previously exposed to an emotion-provoking stimulus other than handling and those subjected to a "gentling" procedure during early life did not differ significantly. Both experimental groups, reacting in much the same manner, evidenced significantly more activity than unmanipulated control animals. Although not significant, a similar trend was noted on an elevated Y maze. It appears, then, that a distributed emotion-provoking experience, as well as handling, is capable of raising, temporarily at least, the threshold for emotional reactivity.

The following interpretations are offered with regard to the conflicting results of previous studies. Using an open field, Stanley and Monkman (1956) found no differences between the defecation and activity of groups of animals given early experience with electric shock and controls which were *picked up and placed into the apparatus* without receiving shock. Such a control group would be comparable to my handled animals and not expected, in light of the above findings, to differ from "emotionalized" animals. The same interpretation might also account for the results of Griffiths and Stringer (1952) whose control animals "received approximately the same treatment as the experimentals except that they were not exposed to the types of intense stimulation mentioned" (p. 302).

Support for the inadequacy of these control groups comes from a study wherein unmanipulated controls were compared with a group receiving electric shock and one placed on the grid without shock (Chevalier & Levine, 1955). Ignored animals proved inferior when given avoidance training at 60 days of age.

The results of the Hall and Whiteman (1951) study, which concluded that intense stimulation in infancy results in emotional instability, may be due to a number of variables. The stimulus situation used was exactly like that employed in eliciting audiogenic seizures (Witt & Hall, 1949), and the strain of mice used was previously shown to have a high proportion of seizure-susceptible animals (Hall, 1947). Moreover, as has been pointed out (Beach & Jaynes, 1954), the open field situation in which these animals were later tested was the same as that in which the early experience occurred.

Each of the variables discussed above (including handling) represents a situation foreign to the accustomed laboratory environment of the animal. As any researcher with animal experience will testify, the rat initially resists attempts at being handled, and Tryon et al. (1941a, 1941b) has even used the handling situation as one reliably eliciting emotional behavior.

While the behavior of handled animals and those subjected to an emotion-provoking situation in early life does not differ on certain tests (Ader, 1957) it would not be justified to define handling as an emotion-provoking stimulus. In the present thesis the position is taken, however, that both the handling regime and emotion-provoking situation constitute, initially at least, a change from the basal level of stimulation. It is hypothesized, then, that handling or any emotion-provoking stimulation of a nontraumatic nature and distributed throughout the *early* life of an individual is capable of raising the threshold for emotional reactivity to subsequent, qualitatively different emotion-provoking situations. Although on different levels, one may note the similarity of this hypothesis to that advanced by Griswold and Gray (1956).

The fact that the particular type of "early" experience in previous studies remained the same on successive occasions suggests that learning may have been involved. The superior learning performance demonstrated by handled animals (Bernstein, 1952; Karn & Porter, 1946) may simply be due to their having adapted to the handling involved in the subsequent experimental procedure. And superior performance in avoidance conditioning with shock as the unconditioned stimulus appears to be due to early experience with electric shock. If adaptation is a significant variable, what might be the effect of qualitatively different experiences during early life, none of which were repeated often enough to allow for any specific adaptive response?

Besides the lesser degree of emotional behavior displayed by handled and "emotionalized" animals (Ader, 1957; Weininger, 1956), handled animals also show less organic damage under severe stress conditions (Weininger, 1954, 1956). One might ask whether animals subjected to emotion-provoking stimulation during infancy would demonstrate the same resistance, and, if so, whether there were qualitative and/or quantitative differences in the physiological mechanisms mediating such resistance. That such differences may exist is suggested by Herrington and Nelbach's (1942) study wherein "disturbed" animals showed an increase in both body weight and weight of the adrenal glands over control animals; the former being characteristic of "gentled" and the latter of nongentled animals (Weininger, 1956).

The purpose of the present research, then, is (*a*) to test the hypothesis that both handling and emotion-provoking stimulation of a nontraumatic nature, if distributed throughout early life (as opposed to manipulation at maturity), are capable of raising the threshold for emotional reactivity to subsequent emotion-provoking situation, and (*b*) to investigate the influence of handling and various types of emotion-provoking stimulation during early life on subsequent resistance to a prolonged stress situation which, if continued, would result in death.

METHOD

The rats used in this investigation were bred in the colony of the Department of Psychology at Cornell. At weaning (22 days of age) 104 albino rats from 11 litters were divided into 10 groups of 10 or 11 animals. Groups were equated for sex and body weight, and, insofar as was possible, at least one animal from each litter went into each group. Once divided, groups were randomly assigned to one of 10 conditions. Five of these groups were to be manipulated during early life (Part A of the experiment) and the remaining five groups, representing the same five conditions, were to be manipulated during maturity (Part B of the experiment). The inclusion of comparable groups of mature animals is essential if one is concerned with the effects of *early* as opposed to *previous* experiences on a given variable.

All animals were maintained in a well illuminated, temperature-controlled room. Four to six rats were housed to a wire mesh cage measuring either 12 × 14 × 8 in. or 12 × 10 × 8 in. An ad libitum food and water schedule was maintained until the final stress measures were taken.

On the day following weaning the differential treatment of animals of Part A was begun and it adhered to the following program for 30 successive days. Weekly weight measures were taken throughout the experiment.

Group H ($N = 10$) was handled for approximately 3 to 4 min. per day in a random, nonstandardized manner. Animals were stroked, scratched, and allowed to wander over the experimenter's hands and arms.

Animals of Group E_D ($N = 10$) were lifted by the tail with a large forceps and placed individually into a cylindrical wire mesh cage measuring 5 in. in diameter and 8 in. in length. These dimensions provided ample room for the largest animal to move about. The cage was suspended from an overhanging beam by two ropes 21 in. long. In this position, animal and cage were lifted and then allowed to swing freely. This procedure was repeated four times each day, a period equal in time to that of the handling procedure.

Group E_V ($N = 10$) was exposed to a variety of emotion-provoking stimuli. These included (*a*) swinging in the above-mentioned cage, (*b*) being thrown into the air through one complete somersault in this same cage, (*c*) being placed under a gallon-sized cardboard liquid container into which small air holes had been punched, (*d*) being rolled across the floor in this same container, (*e*) having their tails pinched with the forceps, (*f*) vigorous rattling of their cage, (*g*) exposure to electric shock, and (*h*) being placed into a metal cooking pot (12.5 in. in diameter and 8 in. high) which *E* covered and banged upon with a stick.

The above situations were chosen from a longer list of noxious stimuli as being least likely to cause physical damage to the animal. Some of these procedures have been employed in previous studies (Griffiths & Stringer, 1952; Herrington & Nelbach, 1942; Higginson, 1930; Humphrey & Marcuse, 1939; Scott, 1955). When applicable, these stimuli were presented for a period equal to that involved in manipulation of the preceding groups. Being tossed into the air in the restricting cage was repeated 10 times on one day. The electric shock, adjusted for each animal to that voltage required to elicit a moderate squeal, was presented intermittently over a 3-min. period, on the grid of a Skinner box. Situations *e* and *f* were carried out while all the animals were together in their living cage. The remaining stimuli were presented to each animal individually. Whenever possible, animals of this group were manipulated by means of a large forceps, or, if by hand, were picked up by the tail. Only one of these stimulus situations occurred on a given day. The order of presentation was random.

Group E_M ($N = 10$) was exposed to the same emotion-provoking stimuli as Group E_V; however, in this case they were all presented on the same evening. The situations immediately followed upon one another, and might well be described as "traumatic." This massed experience occurred at a mean age of 35 days for animals of Part A and at a mean age of 148 days for animals of Part B.

Group C ($N = 11$) served as controls, experiencing no manipulation by the experimenter other than the proximity associated with feeding, the cleaning of cages, and being weighed weekly. Feeding necessitated opening and closing cages each day. Cleaning involved *E*'s being active in the vicinity of the cages every second or third day.

When the animals reached a mean age of 68 days, each cage of rats, in turn, was placed on a table between two other covered cages of equal size. The experimenter stood at a distance and recorded the time required by each animal to emerge from its living cage. The same procedure was repeated at 78 days of age.

At a mean age of 69 days, animals were introduced singly into an open-field situation for 1 min. on 5 successive days. The apparatus was 5 ft. in diameter with 2-ft.-high sheet metal walls. The unpainted Masonite floor was chalked off into four concentric circles and 7.5-in. squares (Weininger, 1956). The field was placed on the floor of a well illuminated room just to the side of a series of fluorescent ceiling lights.

An animal was placed in a square next to the wall and his subsequent activity recorded by a tracing of his path on a scaled diagram of the apparatus. Movement into an adjacent square was recorded when the animal moved all four paws across the boundary line. The same criterion held for movement from one to another of the concentric circles. The total number of squares traversed and movements with respect to the concentric circles were noted.

At a mean age of 75 and 76 days, animals were placed onto one end of an elevated Y maze (Montgomery, 1953). Measuring from the midline of the maze, the 32-in. arms came together within a circle 15 in. in diameter. Measuring from the outer edge, the arms were 25.5 in. long, 7.25 in. wide, with 1-in. high walls. The entire maze was painted flat black. Each arm was chalked off into two 10-in. units and one 12-in. unit reaching into the center. The apparatus rested on three piano stools at a height 3 ft. from the floor.

Half the animals of each group were run on one day. A trial began when the animal moved four paws into the unit adjoining the one into which he was placed. Once started, the animal's position was determined by the unit in which his front paws rested. Recorded were the time required to move from the starting position and the total number of units traversed during each of three successive minutes.

At a mean age of 87 days, 7 animals from each group were weighed and then subjected to 48 hr. of immobilization plus food and water deprivation in a somewhat colder than usual experimental room (20° C.).

Immobilization was accomplished by means of $\frac{1}{8}$-in., 27-gauge wire mesh. A piece approximately 8 in. square was first rolled into a funnel and four or five longitudinal cuts were made at each end. Those at the apex were folded in, closing off the wire into a truncated cone. Those at the base were temporarily bent outwards. A heavy rubber band was placed around the middle. Just past the middle, toward the base, a slit, approximately ½ in. wide was cut. This slit extended around ⅓ to ½ the diameter of the cone at that point. Through this slit a 1-ft.-long insulated wire was inserted so that the ends were on the outside and the middle hugged the circumference of the cone. The cone and wire were held in the left hand and the experimental animal was literally "run" into the wire cage, whereupon *E* closed off the base of the cone locking the animal in place. An excess of room for a given animal was eliminated by pinching in the wire at appropriate places. This procedure allowed for rapid immobilization and prohibited escape.

The electrical wire extending from the immobilization cage was actually two wires joined in the middle by a thin rubber tube packed with powdered graphite. This section partly surrounded the middle of the animal's body. The expansion and contraction of the animal's body at every breath changed the diameter of the graphite-filled tube, thereby changing the current flow through the wire. This change was amplified and caused a kymograph pen to chart the pattern of respiration on a roll of paper moving at a fixed number of revolutions per minute.[7] To facilitate the reading of respiration rate, the kymograph pen also closed a leaf switch activating an automatic cumulative counter.

Samples of the respiration of animals in stress were taken at 15-min. intervals during the first hour and at 8-hr. intervals (plus or minus 15 min.) thereafter. The availability of only one amplifier and pen as well as the time required in the later sacrificing of animals necessitated placing the animals in stress at hourly intervals. Two to five animals were immobilized during any one morning, afternoon, or evening period.

Once immobilized, an animal was placed, stomach down, upon a long Masonite board (9 × 32 in.) in a small experimental room opposite that in which the recording devices were located. Screws protruding from the board kept the immobilization cage from rolling over. Each "slot" had its own "jacks" into which the respiration leads could be inserted. All of these were wired into a separate control panel so that the experimenter could change the animal being recorded without having to enter the experimental room. Occasionally, certain 8-hr. records for a given animal were omitted for practical reasons; at other times mechanical failures contributed to the smaller *N* representing the respiration measure after a given number of hours in stress.

At the end of the 48-hr. immobilization period, animals were removed from the wire mesh and weighed. One cc. of nembutal (approximately 6 mg./kg. of body weight) was injected intraperitoneally, and when the animal was completely anesthetized its throat was cut. The heart, liver, kidneys, spleen and adrenal glands were removed, placed into a 10% solution of formaldehyde, and weighed during a period one to two weeks later. A separate, coded jar was used for each animal so that *E* had no knowledge of the group to which an animal belonged at the time of weighing.

Three randomly chosen animals from each group (four in the case of Group C) served as stress controls. These animals were sacrificed and autopsied without having been subjected to the immobilization procedure.

Animals in Part B of the experiment were subjected to the same conditions as described above beginning at a mean age of 136 days. Their early experience was the same as that described for the control group of Part A.

On the basis of certain incidental observations and a perusal of the stress data of Part A, certain additional procedures were employed. Hourly respiration measures were taken during the first 8 hr. of immobilization. Immediately upon sacrificing the animal, samples of blood were drawn from the heart and the procedure outlined by Loewe, Rosenblatt, and Hirsch (1946) was employed to determine the time required for clotting to occur. Finally, a human oral thermometer was used to take aperiodic samples of rectal temperature during the 48-hr. period of immobilization.

An outline of the experimental procedure appears in Table 1.

RESULTS

Body Weight

Following Scott's (1955) suggestion, the data from the weekly weight measurements were analyzed in terms of the percent increase in body weight rather than the absolute weight values of each group. By converting the data in this manner, one compensated for any possible tendency on the part of a given group to be initially heavier. The growth curves for animals of Part A were not amenable to clear graphic presentation. The five curves were virtually superimposed upon one another. For animals of Part B, the mean percentage increase in body weight over the initial body

---

[7] This method of obtaining respiration measures was modeled after that devised by A. U. Moore who developed the technique for large animals at the Cornell Behavior Farm. The author is indebted to Dr. Moore for the use of his equipment and his assistance in setting up the present apparatus.

TABLE 1

Outline of Experimental Procedure

Age (in days)		Group				
Part A	Part B	H	E_D	E_V	E_M	C
22	22	Weaning (all groups)				
23–53	136–166	Handling	Single, distributed emotion-provoking situation	Variety of emotion-provoking situations, distributed	Variety of emotion-provoking situations, massed (at 35 or 148 days of age)	No manipulation
54–67	167–180	No manipulation				
68	181	Emergence from cage (I)				
69–73	182–186	Open field				
74	187	No manipulation				
75, 76	188, 189	Y Maze				
77	190	No manipulation				
78	191	Emergence from cage (II)				
79–86	192–199	No manipulation				
87	200	Stress NS*				
		Sacrifice all animals				

* Stress control animals underwent no manipulation during this period.

weight has been plotted in Fig. 1. While young animals doubled their body weight in the two weeks of manipulation following weaning, animals of Part B, beginning manipulation at 136 days of age, did not gain 12% of their premanipulation weight throughout the entire period of measurement. Despite this difference in the magnitude of the values, and in order to uncover any interaction effect should one exist, one analysis of variance was applied to all of the data (Table A).[8] As was to be expected, the difference between the weight gains of early and late manipulated animals is significant.

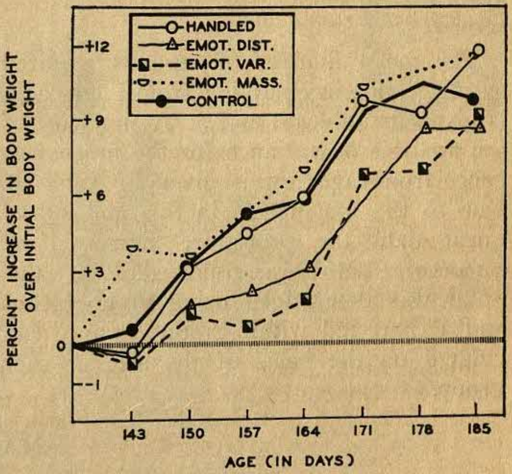


Fig. 1. Mean percentage increase in body weight over initial body weight for animals manipulated late in life.

The *F* test did not reveal significant weight differences as a function of the differential treatment of the various groups. It is of interest, however, that at the beginning of the manipulation of the late groups, both handled animals and those exposed to

---

[8] Tables summarizing the analysis of variance for this and subsequent data have been designated alphabetically and have been deposited with the American Documentation Institute. Order Document No. 5819 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 2

NUMBER OF SECONDS REQUIRED FOR EMERGENCE FROM CAGE
(TEST I)

Period	Group H		Group E_D		Group E_V		Group E_M		Group C		Combined Mean
	N	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*	Mean	
Early	10	303.5	10	332.5	10	449.4	10	479.8	11	472.3	407.5
Late	9	521.6	9	711.9	9	670.2	8	696.5	9	458.5	611.74
Combined		412.5		522.2		559.8		588.1		465.4	

distributed emotion-provoking stimulation showed a loss in body weight. The remaining two groups which experienced no manipulation at this time (C and E_M) showed the expected increase in weight. During the week in which Group E_M was subjected to the massed emotion-provoking stimulation (148 days) it, too, showed a decrease in weight. While consistent with the present hypothesis, not too much emphasis should be placed upon such data because of the seemingly random fluctuations in weight shown by all of the other groups at some time during the 7-week period.

Emergence From Cage

The mean number of seconds required by all groups in emerging from their cages is given in Tables 2 and 3. A summary of the analysis of variance for the first emergence-from-cage test is given in Table B. Due to the unequal *N* in this and subsequent data the procedure suggested by Snedecor (1946) was followed. The value of *F* was determined using group means and a corrected within-group variance calculated on the basis of the size of each group.

The analysis showed only the interaction between the type of experience and the time at which it occurred (early or late) to be significant ($F = 28.79$, $df = 4$, $p < .01$). In order to determine which groups were responsible for this significant *F*, independent *t* tests were run between early and late animals for the various experimental conditions. While no difference was demonstrated to exist between early and late control animals, all animals subjected to manipulation late in life required significantly more time before leaving their cages than their early manipulated counterparts ($p <$ .01 in all cases).

Bartlett's test for homogeneity of variance was applied to all data before performing an analysis of variance. Unless otherwise specified, the value of chi square was not significant. In the case of the second emergence-from-cage test, Bartlett's test revealed heterogeneity of variance ($\chi^2 = 22.02$). A square-root transformation was applied and served to normalize the data as indicated by reapplication of the Bartlett test. A summary of the analysis of variance on the transformed data appears in Table C. Here again there

TABLE 3

NUMBER OF SECONDS REQUIRED FOR EMERGENCE FROM CAGE
(TEST II)

Period	Group H		Group E_D		Group E_V		Group E_M		Group C		Combined Mean
	N	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*	Mean	
Early	10	449.0	10	159.7	10	124.7	10	334.7	11	220.4	257.7
Late	9	605.5	8	494.6	9	505.5	7	549.0	8	501.0	531.12
Combined		527.2		327.1		315.1		441.8		360.7	

TABLE 4

Number of Squares Traversed Per 5 Trials in the Open Field

Period	Group H		Group E_D		Group E_V		Group E_M		Group C		Combined Mean
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	
Early	10	34.60	10	27.34	10	33.60	10	27.90	11	26.67	30.02
Late	9	20.08	9	18.99	9	13.82	8	12.37	9	16.64	16.38
Combined		27.34		23.16		23.71		20.13		21.65	

were no differences in the time required for animals to emerge from their cages as a function of the type of manipulation experienced, and unlike Test I, no interaction effect was demonstrated. Animals manipulated later in life, however, do require more time in emerging from the cage ($F = 21.35$, $df = 1$, $p < .01$). The mean number of seconds required by early manipulated animals was 257.7 as opposed to 531.1 sec. for animals manipulated later in life. Although not significant in the case of control animals, this same result was observed in the case of manipulated groups in the first test.

Open Field

Activity in the open field in terms of the mean number of squares traversed for both early and late groups is shown in Table 4, and Table D summarizes the analysis of variance for the activity measure. The only significant value ($F = 25.61$, $df = 1$, $p < .01$) occurs between the groups of animals manipulated early (with an over-all mean of 30.02 squares traversed) and those groups manipulated later in life (with an over-all mean of 16.38).

The data with respect to the concentric circles of the open field are given in Table E. In the analysis by Weininger (1956), the percentage of times that animals were in each concentric circle of the field was plotted. The present data are plotted in terms of the number of entries made by the animals of each group into Circles 2, 3, and 4 of the open field (Fig. 2). This latter mode of presentation was chosen because of the rat's tendency to remain close to the walls and avoid the lighter, open center of any given field. The great majority of activity in the open field occurred in the outermost circle (Circle 1). An animal entering inner circles of the field would invariably return within a few seconds to the outermost circle, usually after having "passed through" rather than having "wandered about" the inner areas. In returning to the outer circle, the animal must again pass through inner circles of the field, thereby increasing differentially the amount of activity in Circles 2, 3 and 4, and in that order. It is felt that a somewhat more accurate account of the distance away from

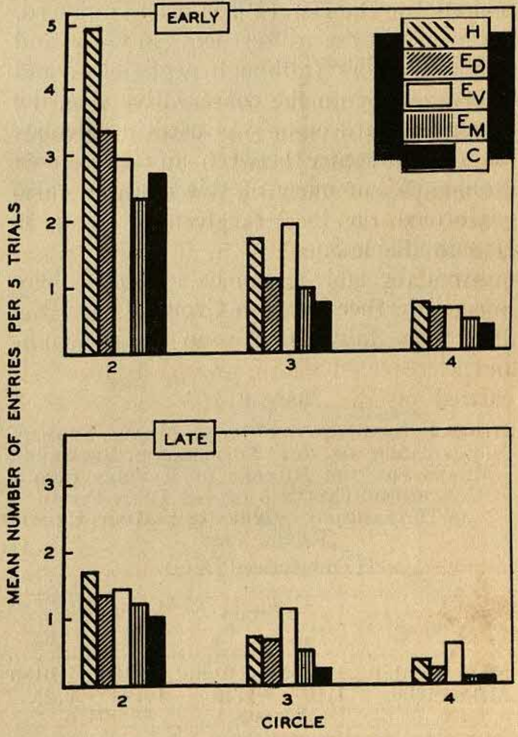


Fig. 2. Mean number of entries into concentric Circles 2, 3, and 4 of the open field.

the walls of the apparatus (as a measure of "emotionality") is achieved by counting only the number of forward-going entries into the inner circles.

For concentric Circles 2 and 3, the analyses of variance are summarized in Tables F and G. In view of the many zero scores, 0.5 was added to each score and the square root of this value was taken. The analyses were performed on the transformed data.

Animals tested early in life show significantly more entries into Circles 2 and 3 than animals tested at maturity ($F = 15.00$ and $F = 41.82$, $df = 1$, $p < .01$ in both cases). No interaction effect was demonstrated in either of these circles. For Circle 2, the value of F revealed no differences resulting from differential experience, while the value of F (3.64, $df = 4$) for the variable of experience in Circle 3 was significant at the .01 level.

In order to determine which of the differences among the various types of experience may be considered significant and which may not, the *new multiple range test* devised by Duncan (1955) was employed. Because of the differences in N's and standard errors (although negligible), and in order to be on the conservative side, the larger standard error was used in calculating the difference between any two given groups. A summary of this analysis (also performed on the transformed data) is given in Table 5.

From the table it can be seen that there is no difference between Groups C and E_M. The mean value for Group E_D is significantly greater than that for either C or E_M, and, although Groups H and E_V do not differ, both show significantly more entries into concentric Circle 3 than do all the other groups.

TABLE 5

Standard Error of the Means, Means, and the Significance of the Difference Between Means for the Number of Entries Into Concentric Circle 3 of the Open Field as Determined by the New Multiple Range Test

(Transformed Data)

	Group				
	C	E_M	E_D	H	E_V
$SE_{\bar{x}}$	.0170	.0185	.0180	.0180	.0180
Mean	1.00	1.01	1.10	1.19	1.25

Note.—Any two means that are *not underscored* by the same line are significantly different at the .05 level of confidence. Means underscored by the same line are not significantly different.

TABLE 6

Defecation Measures Over 5 Trials in the Open Field for Animals of Part B

Measure	Group				
	H (N=9)	E_D (N=9)	E_V (N=9)	E_M (N=8)	C (N=9)
Mean	.09	.18	1.28	1.57	.75
No. animals defecating	1	1	4	7	5
% animals defecating	11.1	11.1	44.4	87.5	55.5

With regard to the innermost circle, no statistical tests were employed since the majority of scores were zero. That is, very few animals in any group, whether tested early or late, entered Circle 4. Suffice it to point out that the pattern set down for Circles 2 and 3 (Groups H, E_D, and E_V having greater mean values than Groups E_M and C) is repeated for Circle 4 (Fig. 2).

Defecation was also studied as a variable in the open-field situation. In the case of those groups manipulated in early life, no more than one animal in each group defecated. Animals manipulated later in life showed a sufficient amount of defecation to warrant statistical analysis. Defecation measures for animals of Part B of the experiment are given in Table 6. Because of the preponderance of zero scores and the rather wide range of scores it was felt that any assumption about the normality of the data would be unrealistic. Since defecation data from Part A was not being analyzed concomitantly, the use of a nonparametric technique in the present instance did not involve the loss of any interaction effect. For these reasons the data were not transformed, and the Kruskal-Wallis one-way analysis of variance (Kruskal, 1952) was employed. The obtained value of H (8.842) permits rejection of the null hypothesis at the .069 level of confidence. That is, these

TABLE 7

LATENCY IN Y MAZE

(In Seconds)

Period	Group H		Group E_D		Group E_V		Group E_M		Group C		Combined Mean
	N	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*	Mean	
Early	9	38.1	9	42.8	10	15.3	10	21.5	11	19.0	27.3
Late	9	38.1	7	58.3	8	147.8	7	99.0	7	143.6	97.4
Combined		38.1		50.5		81.5		60.2		81.3	

results could have been obtained by chance about 7 times in 100.

Using the Mann-Whitney (1947) *U* test and the extended tables given by Siegel (1956), the following independent comparisons consonant with the present hypothesis were made:

Comparing Group E_M with C, the critical value for *U* at the .05 level is 15. The obtained value of *U* (22.5) does not allow rejection of the null hypothesis, the difference between these groups being no greater than would be expected on the basis of chance alone.

Comparing Group E_D with Group E_V, the critical value for *U* at the .05 level is 17. Here again, the observed value of *U* (26.0) does not allow rejection of the null hypothesis.

Comparing Group H with Groups E_D and E_V, the critical value for *U* at the .05 level is 99. The observed value of *U* is 65.0, indicating that Groups E_D and E_V, combined, excrete significantly more fecal boluses in the open field than do handled animals.

Comparison of Groups H, E_D, and E_V as a whole with Groups E_M and C (pooled) provides a population large enough so that the distribution of *U* closely approximates that of a normal distribution. For this comparison $Z = 3.77$ ($p < .01$). Nonmanipulated control animals and those animals subjected to the massed emotion-provoking stimulation, combined, show more defecation in the open field than the remaining three groups taken together.

Y *Maze*

As a correction for the obtained heterogeneity of variance ($\chi^2 = 161.1$), a square-root transformation was applied to the data of the measure of latency in the Y maze. The mean latencies in the original data appear in Table 7. A summary of the analysis of variance on the normalized, transformed data is given in Table H.

The only value of *F* which is significant at better than the .05 level of confidence occurs between the groups of animals manipulated early (with a mean of 27.3 sec.) and those manipulated later in life (with a mean of 97.4 sec.) ($F = 9.64$, $df = 1$, $p < .01$).

Activity in the Y maze, measured by the number of units traversed, is plotted in Fig. 3 and 4. A summary of the analysis of variance appears in Table I. In this case the variable of experience (the various types of manipulation) yields a significant

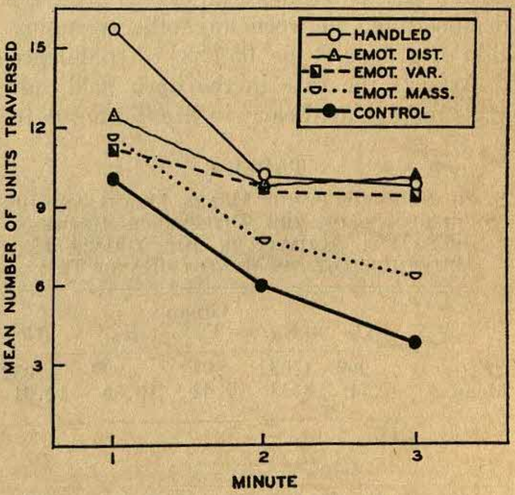


FIG. 3. Activity per minute in the Y maze for all groups of Part A.

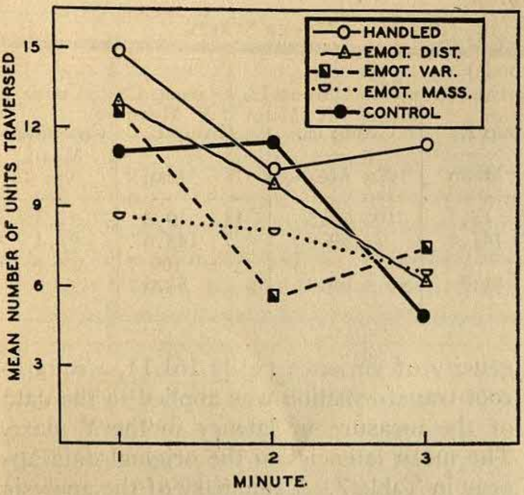


FIG. 4. Activity per minute in the Y maze for all groups of Part B.

value ($F = 3.22$, $df = 4$, $p < .05$). As in the case of concentric Circle 3 of the open field, the *multiple range test* was used to determine which of the differences among the various types of experience may be considered significant at a satisfactory level of confidence (.05). A summary of this analysis is given in Table 8.

From the table it can be seen that, while no difference has been demonstrated between Groups C and E_M, both show significantly less activity than handled animals. Group C is also significantly less active than Group E_D. Differences in activity were not demonstrated between any other combination of groups at the .05 level of confidence.

As was the case in the open field, animals manipulated early in life did not defecate in the Y maze in sufficient quantity or in number of animals defecating to warrant statistical analysis. Although the amount of defecation was not great for animals manipulated later in life, the data were analyzed in the same way as for the open field. The Kruskal-Wallis one-way analysis of variance was employed. The obtained H (0.35) failed to approach a satisfactory level of confidence.

TABLE 8

STANDARD ERROR OF THE MEANS, MEANS, AND THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS FOR ACTIVITY IN THE Y-MAZE AS DETERMINED BY THE MULTIPLE RANGE TEST

	Group				
	C	E_M	E_V	E_D	H
$SE_{\bar{x}}$	.969	1.032	.942	.999	1.032
Mean	7.71	8.33	9.42	10.86	12.01

Note.—Any two means that are not underscored by the same line are significantly different at the .05 level of confidence. Means underscored by the same line are not significantly different.

Respiration Rate Under Stress

The first sample of respiration rate was taken as soon as an animal was immobilized. For animals of Part A, samples were taken during the first hour of stress, at hourly intervals during the first 8 hr. of stress, and at 8-hr. intervals thereafter. In this latter case, additional samples were taken whenever time allowed, although it was the first 8 hr. which were concentrated upon. These additional measures were taken in order to determine if the changes in respiration over the 48-hr. period of stress were similar in pattern to changes characteristic of Selye's (1950) *general adaptation syndrome,* as was indicated by the similar form of the respiration patterns for a few animals of Part A.

In Table 9 the mean number of breaths per minute is recorded for the initial measure and the period just before being removed from the immobilizing wire. The analyses of variance (Tables J and K) indicated no differences to exist between any groups.

The data from respiration rates were not readily amenable to either concise presentation or statistical analysis. In many cases mechanical difficulties resulted in a smaller N representing the rate of respiration after a given number of hours in stress. The number of animals for whom a complete record was obtained was relatively small. Under these circumstances the graphic representation of the mean respiration rate through the 48 hr. of stress would hardly be representative. Moreover, even if records were available for the total N at each period of measurement, the variability in the magnitude and the time relations of the

TABLE 9

INITIAL AND FINAL RESPIRATION RATES (PER MIN.) UNDER STRESS

	Group H		Group E_D		Group E_V		Group E_M		Group C		Comb.
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	
						Early					
Initial Respiration Rate	7	111.0	6	107.5	5	79.0	7	89.2	6	104.8	98.30
Respiration rate after 48 hr. of stress	7	73.2	7	79.8	2	92.0	6	70.8	5	65.0	76.2
						Late					
Initial Respiration Rate	6	114.0	6	101.7	4	104.2	5	104.2	6	109.7	106.8
Respiration rate after 48 hr. of stress	4	76.0	4	63.0	4	59.5	5	63.0	4	78.7	68.0

changes noted in certain animals could not permit averaging. It is quite possible that by obtaining measures of central tendency the over-all pattern would be misrepresented or eliminated.

In some cases such difficulties may be eliminated by treating the data in terms of the amount of deviation from some basal level. This was not possible with the present apparatus, and the author knows of no method whereby respiration rate may be obtained on the rat under "normal" conditions. The initial measure in this situation followed immobilization, and, as would be expected, animals showed a highly elevated rate of respiration in response to the stress. It should be emphasized, therefore, that while no differences were demonstrated to exist between the initial level of respiration of the various groups in response to stress, the more desirable, and perhaps more sensitive, measure of the change from a resting state could not be applied to the data. The same is true for the measure of respiration after 48 hr. of stress.

The amount of available data on respiration was not sufficient to permit any kind of evaluation of the shape of the individual curves in terms of the *general adaptation syndrome*. Sample records are given in Fig. 6 and are discussed in conjunction with the measure of body temperature.

Body Temperature Under Stress

The possible involvement of the *general adaptation syndrome* under the conditions of stress imposed in the present study also provoked the measurement of body temperature. According to Selye (1950) hypothermia is characteristic of the *alarm reaction* (the initial response to stress). Similarly, other investigators (Bartlett, Bohr, & Helmendach, 1954; Bartlett, Bohr, Helmendach, Foster, & Miller, 1954; Grant, 1950) have demonstrated that the stress of restraint results in a hypothermic response. Of particular interest is the report that this hypothermia is more pronounced in the more emotional animals (Bartlett, Bohr, Helmendach, Foster, & Miller, 1954).

As with respiration, the first measure of body temperature was taken immediately following immobilization. While the physical manipulation involved in this process would certainly remove an animal from its resting state, it was assumed that a change in temperature would not come about as rapidly as a change in respiration. The change in temperature from this initial level is plotted in Fig. 5 for the first 8 hr. of stress.

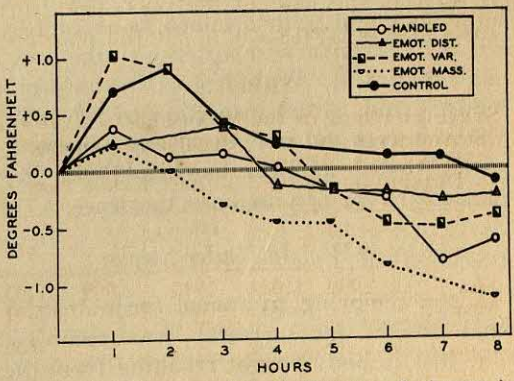


FIG. 5. Mean number of degrees of change in body temperature from initial body temperature during the first 8 hr. of stress.

Contrary to previous reports (Bartlett, Bohr, & Helmendach, 1954; Bartlett, Bohr, Helmendach, Foster, & Miller, 1954; Grant, 1950) the initial response to the immobilization was a hyperthermic one. For individual animals this hyperthermia preceding a hypothermic response lasted from 1 to 14 hr. Only 7 out of 26 animals showed a decrease in body temperature when the first measure (½–1 hr. after the initial measure) was taken, and these seven were dispersed in the five groups.

An *F* test revealed no differences in the mean change in body temperature from the original level between any of the groups during the first 8 hr. of stress. A summary of the analysis appears in Table L. In Table M is a summary of the analysis of body temperature taken after 48 hr. of stress. Again, the calculations were in terms of the degrees of change from the initial level, irrespective of the direction, and, again, the value of *F* was not significant.

Sample records of the body temperature and respiration rate of individual animals are shown in Fig. 6. These particular animals were chosen as representative of the variability in the magnitude and the time relations of the changes which occurred. They also illustrate the similarity in the course taken by both body temperature and respiration rate during a 48-hr. period of stress. Whether curves such as these may be interpreted in terms of Selye's *general adaptation syndrome,* and the problem of how this might be determined, is of interest but is of secondary importance to the present experiment. With respect to the problem at hand, visual inspection of the individual curves did not reveal any consistent pattern which might be associated with any particular type of previous experience.

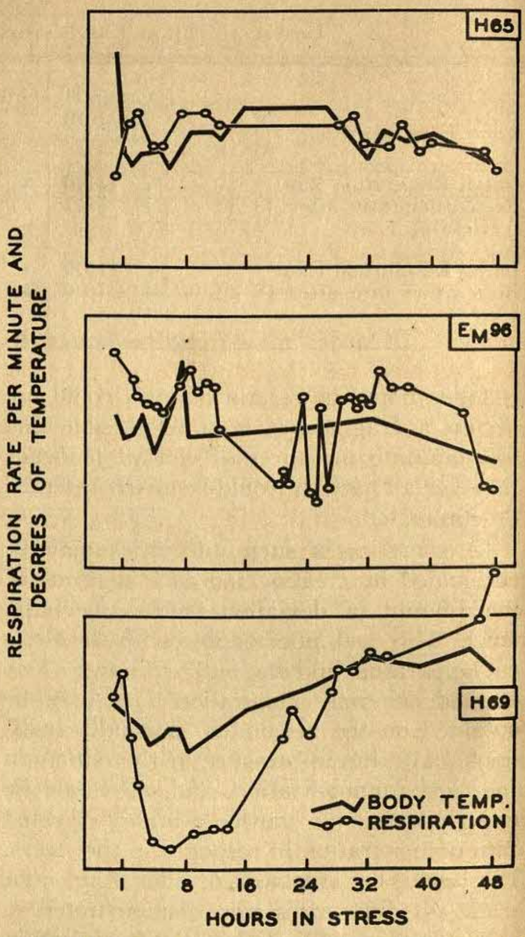


FIG. 6. Sample graphs of respiration rate and body temperature over 48 hr. of stress.

Loss in Body Weight Under Stress

Upon removing an animal from stress, a final weight measurement was recorded. The loss in body weight resulting from the immobilization was calculated in terms of the percentage of the prestress weight of the animal. The mean values for each group are given in Table 10. Since the *F* test is not applicable to percentage data with an upper limit, an arcsin transformation was applied to the data (Snedecor, 1946). A summary of the analysis of variance on the transformed data appears in Table N. The value of *F* was not significant.

Blood Clotting

While in the process of removing certain organs from the animals of Part A of the experiment, it was noted that the pools of blood which collected were sometimes clotted and sometimes not. It appeared, moreover, that the blood of nonstressed animals clotted faster than that of stressed

TABLE 10

PERCENTAGE OF LOSS IN BODY WEIGHT DUE TO STRESS

Period	Group H		Group E_D		Group E_V		Group E_M		Group C		Comb.
	N	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*	Mean	
Early	7	16.60	7	15.53	6	15.58	7	17.80	7	17.56	16.61
Late	6	15.35	6	14.68	6	14.73	5	17.78	5	15.84	15.68
Combined		15.97		15.10		15.16		17.79		16.70	

animals. In order to verify these original observations, the clotting time of animals of Part B was measured by a standardized method of analysis. The average blood coagulation time of the rat is 2.5 min. (Creskoff, Fitz-Hugh, & Farris, 1949). The animals in this study showed a range of from approximately 1.0 to approximately 5.0 min. The median clotting time was 3.3 and 2.0 min. for combined stressed and nonstressed animals, respectively.

The method involved in determining the coagulation time allowed only an ordinal scale of measurement. The median clotting times and the number of animals with clotting times above and below the over-all median of 3.0 min. are given in Table 11. Comparing the blood of all stressed with all nonstressed animals one finds that in 20 of 26 stressed animals 3 min. or more elapsed before coagulation took place. The blood of only 2 of 12 nonstressed animals required this much time. The probability that this difference is due to chance is .007 as determined by the Mann-Whitney *U* test.

A comparison of the several stressed groups using the Kruskal-Wallis one-way analysis of variance resulted in an *H* of 11.095 which is significant at the .05 level. Using the Mann-Whitney *U* test, independent comparisons were then made. The difference between Group E_M with a median of 2.75 min. and Group C with a median of 3.75 min. is significant ($U = O$, $p = .001$). The difference between Groups E_D and E_V and that between Group H and Groups E_D and E_V, taken together, were not significant at the .05 level. The only remaining independent comparison (Groups H, E_D, and E_V against Groups E_M and C) was also not significant at the .05 level. While it is difficult to draw conclusions on the basis of a limited number of independent comparisons, the similarity between the blood coagulation time of animals given the massed emotion-provoking stimulation (Group E_M) and nonstressed animals, and the relation of these to the remaining stressed groups, is worthy of note.

TABLE 11

BLOOD CLOTTING IN STRESSED AND NONSTRESSED ANIMALS

(Part B)

Measure	Stressed						Nonstressed
	Groups					Combined	Combined
	H	E_D	E_V	E_M	C		
N	6	6	4	5	5	26	12
No. animals with clotting time ≥ 3 min.	6	5	3	1	5	20	2
Median	3.25	3.40	3.25	2.75	3.75	3.3	2.0

Organ Weights

In Table 12 are given the organ weights for each of the stressed groups and the weights of stressed animals taken together. The organ weights of nonstressed animals have also been combined since in no case was there any indication of a difference between nonstressed animals as a function of experience. The mean values are given in grams per 100 grams of the prestress body weight.

Except in the case of the adrenal glands, the organs of early and/or late nonstressed animals differed in weight from stressed animals. The difference in size of all of the organs (roughly corresponding to the

Bureau Ednl. Psy. Research
DAVID HA[illegible] T[illegible]AINING COLLEGE
Dated ......
Acces No ......

TABLE 12

ORGAN WEIGHTS (GM./100 GM. OF BODY WEIGHT) FOR ALL GROUPS

Organ	Period	Stressed											Non-stressed	
		Group H		Group E_D		Group E_V		Group E_M		Group C		Combined	Combined	
		N	Mean	N	Mean	N	Mean	N	Mean	N	Mean		N	Mean
Heart	Early	7	.282	7	.277	6	.287	7	.265	7	.265	.275	15	.313
	Late	6	.278	6	.296	6	.283	5	.344	6	.301	.299	12	.313
	Combined		.280		.286		.285		.304		.283	.287		.313
Liver	Early	7	2.893	7	2.922	6	2.778	7	2.624	7	2.928	2.830	15	4.090
	Late	6	2.778	6	2.688	6	2.553	5	2.366	6	2.809	2.648	12	3.594
	Combined		2.835		2.805		2.665		2.495		2.868	2.739		3.842
Kidneys	Early	7	.679	7	.615	6	.613	7	.578	7	.634	.624	15	.850
	Late	6	.655	7	.753	6	.696	5	.696	6	.709	.698	12	.837
	Combined		.667		.674		.654		.637		.671	.661		.843
Spleen	Early	7	.127	7	.124	6	.159	7	.119	7	.136	.132	15	.314
	Late	6	.173	6	.205	6	.142	5	.133	6	.187	.169	12	.257
	Combined		.150		.164		.150		.126		.161	.151		.285
Adrenals	Early	7	.0106	7	.0124	6	.0121	7	.0120	7	.0127	.0120	15	.0180
	Late	6	.0152	6	.0139	6	.0135	5	.0130	6	.0161	.0144	12	.0115
	Combined		.0129		.0131		.0128		.0125		.0144	.0132		.0147

differences in weight) are pictured in Fig. 7. These organs were removed from a stressed and nonstressed control animal of Part A of the experiment. The prestress body weight of these animals was exactly the same.

Heart.—In Table O a summary of the analysis of variance for the weight of the heart is given. The value of F for the difference between stressed and nonstressed animals with means of .287 and .313, respectively, is significant at the .05 level ($F = 6.70$, $df = 1$). The only other significant value of F is that for the interaction of the type of experience and the time (early or late in life) at which it occurred ($F = 2.55$, $df = 4$, $p < .05$). In order to determine which groups were responsible for this significant F, independent t tests were run between early and late animals having received the same type of experience. The difference between early and late animals having received emotion-provoking stimulation, regardless of the type or distribution, was significant. In the case of Groups E_D and E_M, late manipulated animals showed a greater weight of the heart than animals manipulated early in life ($p < .01$ in both instances). In the case of Group E_V early manipulated animals showed the greater weight ($p < .05$). No differences were demonstrated to exist between early and late groups of handled or control animals.

Liver.—A summary of the analysis of variance for the weight of the liver appears in Table P. As in the case of the heart, the 48 hr. of immobilization plus food and water deprivation resulted in a loss in weight of the liver ($F = 168.45$, $df = 1$, $p < .01$). For stressed animals the mean was 2.739 gm. and for nonstressed animals it was 3.842 gm./100 gm. of body weight.

The F value for the difference in the weight of the liver between early and late manipulated animals was also significant ($F = 15.32$, $df = 1$, $p < .01$), with groups manipulated early in life (with a mean of 3.22 as compared to a mean of 2.90 for late manipulated animals) showing the greater mean weight. None of the possible interaction effects yielded a significant value of F.

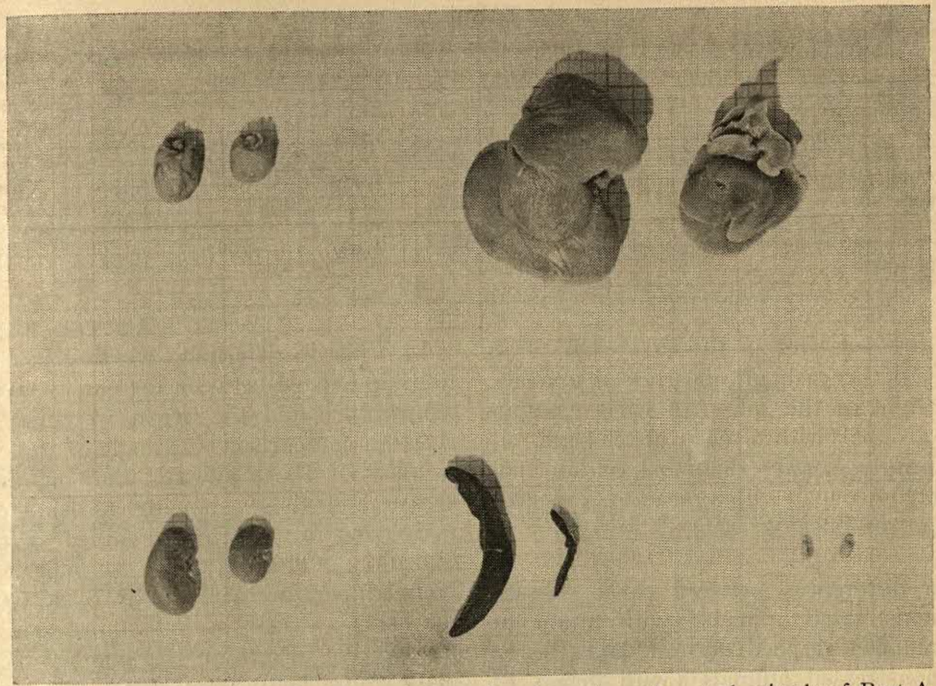


FIG. 7. Photograph of the organs of stressed and nonstressed control animals of Part A. The organ of the stressed animal appears on the right in each instance. Reading from top left to bottom right the organs are the heart, liver, kidney, spleen, and adrenal glands.

Kidneys.—Differences in the weight of the kidneys between stressed and nonstressed animals and between the various experience groups varied according to whether animals were manipulated early or late in life (Table Q).

While a t test did not reveal any difference between early and late nonstressed animals, the kidneys of animals stressed early in life (with a mean of .624 gm./100 gm. of body weight) weighed significantly less than those of animals stressed later in life (with a mean of .698 gm./100 gm. of body weight) ($t = 4.32$, $p < .01$). The differences between early stressed and nonstressed animals (with means of .624 and .850, respectively) and late stressed and nonstressed animals (with means of .698 and .836, respectively) were significant ($p < .01$ in both cases). Furthermore, the significant value of F for the interaction implies that the difference between early stressed and nonstressed animals is greater than that between late stressed and nonstressed animals ($F = 3.46$, $df = 1$, $p < .05$).

The F value for the interaction of experience and the time at which it occurred (early or late in life) was also significant ($F = 3.18$, $df = 4$, $p < .05$). The difference between early and late animals given the same emotion-provoking stimulus throughout the period of manipulation (early and late animals of Group E_D) was significant ($t = 5.88$, $p < .01$). No other differences between early and late animals having undergone a given type of experience were revealed to be significant.

Spleen.—Here again, the significant values of F (Table R) indicate that differences in the weight of the spleen between stressed and nonstressed animals and between the various types of experience varied according to whether animals were manipulated early or late in life.

With a mean of .132 gm. as compared to a mean of .169 gm./100 gm. of body weight the spleen of animals stressed early in life weighs significantly *less* than that of ani-

TABLE 13

ORGAN WEIGHTS (GM./100 GM. OF BODY WEIGHT) OF ANIMALS WHOSE FINAL BODY TEMPERATURE WAS HIGHER AND LOWER THAN THE INITIAL TEMPERATURE

Measure	N	Heart	Liver	Kidneys	Spleen	Adrenals
Final temperature higher than initial temperature	8	.308	2.715	.690	.185	.0153
Final temperature lower than initial temperature	15	.294	2.607	.709	.156	.0140

mals stressed later in life ($t = 4.47$, $p < .01$). In the case of nonstressed animals, the weight of the spleen of early manipulated groups (with a mean of .314 gm.) is significantly *heavier* than that of late manipulated groups (with a mean of .257 gm./100 gm. of body weight) ($t = 4.39$, $p < .01$).

The differences between early stressed and nonstressed animals (with means of .132 and .314, respectively) and late stressed and nonstressed animals (with means of .169 and .257, respectively) were significant ($p < .01$ in both instances). As in the case of the kidneys, the difference between early stressed and nonstressed animals is greater than that between late stressed and nonstressed animals as indicated by the significant value of F for the interaction ($F = 8.24$, $df = 1$, $p < .01$).

As for the interaction of experience and the time at which it occurred ($F = 4.44$, $df = 4$, $p < .01$) no differences were demonstrated to exist between early and late animals of Groups H, E_M, or C. Comparing early and late manipulated animals of Group E_D (a mean of .164 as compared to a mean of .234) resulted in a t of 4.49 which is significant at the .01 level. In the case of Group E_V, animals manipulated early in life show the greater weight of the spleen with a mean of .235 as compared to a mean of .160 gm./100 gm. of body weight for late manipulated animals ($t = 4.52$, $p < .01$).

Adrenals.—A summary of the analysis of variance for the weight of the adrenal glands appears in Table S. The value of F failed to reveal differences under any conditions.

In order to determine whether any relationship existed between the change in body temperature and the organ weights under stress, a comparison was made of the organ weights of those animals showing an increase and those showing a decrease in body temperature after the 48 hr. of immobilization. Omitting those animals whose body temperature after 48 hr. of stress was the same as their initial temperature, and irrespective of group, the number of animals whose body temperature was greater than their initial temperature was 8, and the number showing a decrease in temperature was 15. The organ weights of these two groups are given in Table 13. With the exception of the kidneys, the weights of the organs of animals whose body temperature increased over the initial level tended to be greater than those of animals showing a decrease in body temperature. In no case, however, did t tests reveal any significant differences between the groups.

Relative Performance of the Several Groups

Except in the case of activity in the Y maze, defecation in the open field, and the number of entries into an inner circle of the open field, no differences were demonstrated to exist between the several groups subjected to differential treatments. Despite this lack of statistical significance on a given measure, the relative position of the groups with respect to one another was observed to be fairly constant from one test situation to another. One may inquire, then, concerning the probability that a given group would show a score indicative of greater emotionality (with respect to some other

TABLE 14

RANKINGS OF ALL GROUPS ON THOSE MEASURES CORRELATED WITH EMOTIONALITY

Measure	Part	Group				
		H	E_D	E_V	E_M	C
Emergence from cage (I)	A	1	2	3	5	4
	B	2	5	3	4	1
Emergence from cage (II)	A	5	2	1	4	3
	B	5	1	4	2	3
Activity in open field	A	1	4	2	3	5
	B	1	2	4	5	3
Open field						
Concentric Circle 2	A	1	2	3	5	4
	B	1	3	2	4	5
Concentric Circle 3	A	2	3	1	4	5
	B	2	3	1	4	5
Concentric Circle 4	A	1	2.5	2.5	4	5
	B	2	3	1	4.5	4.5
Defecation in the open field						
Mean	B	1	2	4	5	3
N		1.5	1.5	3	5	4
Latency in the Y maze	A	4	5	1	3	2
	B	1	2	5	3	4
Activity in the Y maze	A	1	3	2	4	5
	B	1	2	4	5	3
Defecation in the Y maze						
Mean	B	1	2	3	5	4
N		1	2	4	5	3
Weight of adrenals (stressed animals)	A	5	2	3	4	1
	B	2	3	4	5	1
Sum of the Ranks		42.5	57.0	60.5	92.5	77.5

group) over a series of tests even though individual tests of the series failed to discriminate between these groups.

Each of the five groups were ranked on all of the tests employed in the experiment. In Table 14 appear the rankings on those measures which are considered indices of emotionality. The lower numbers refer to the less emotional groups. In order to be more conservative, and because they do represent different tests in the sense that they were performed on different animals, Parts A and B were ranked separately. Wilcoxon's (1941) rank test for the comparison of several treatments was used in a preliminary analysis of the data. The value of χ^2r was 27.39 which is significant beyond the .001 level of confidence. Applying Wilcoxon's signed-ranks test (Siegel, 1956; Wilcoxon, 1941) to individual pairs resulted in the following:

No differences were demonstrated to exist between Groups E_M and C. Group E_M is significantly different from the other manipulated groups in the direction of greater emotionality ($p < .01$). Although Group E_V differs from Group E_M, no differences were revealed between Group E_V and control animals on the one hand and the remaining manipulated groups (H and E_D) on the other. While handled animals are somewhat less emotional than animals given the same, distributed emotion-provoking stimulation (Group E_D) ($p < .05$), both of these groups are less emotional than either control animals or those given the massed emotion-provoking stimulation (Group E_M) ($p < .025$).

Discussion

The results from the various tests employed in the experiment support only partially, and with certain qualifications, the hypothesis that both handling and emotion-provoking stimulation of a nontraumatic nature, if distributed throughout early life, are capable of raising the threshold for emotional reactivity. Discussing first the behaviorial situations used to elicit responses indicative of emotionality, as opposed to the physiological reactions to the subsequently imposed stress, one must say that the majority of the measures failed to differentiate between the several groups given different types of manipulation early or late in life. Even so, the relationship of greater or lesser emotionality among the several groups remained fairly constant (and significantly so) over the various test situations, and the direction of the relative position of the groups, in terms of emotionality, does offer some support for the hypothesis. There was no evidence, on the other hand, that, in the situations employed, manipulation early in life resulted in behavior demonstrably different from that effected by such manipulation at a later time.

Except for the second emergence-from-cage test (Parts A and B) and the latency score in the Y maze (Part A only) handled animals appeared to be less emotional than control animals in the various test situations. It had previously been reported (Hunt & Otis, 1955) that handled animals required less time before emerging from their cages; and so they did in the first test of this experiment. In the second test, although the result was not statistically significant, both early and late handled animals required more time than any of the other groups. The lesser emotionality displayed by handled animals relative to nonmanipulated controls was significant in the case of defecation and number of entries into concentric Circle 3 of the open field, and in activity in the Y maze. For those tests which did not yield statistically significant results, the tendency for handled animals to be less emotional than nonmanipulated controls was, as revealed by a comparison of the ranks on the various tests, statistically significant.

That handled animals are less emotional than controls is consistent with Weininger's (1956) results and other previous studies (Ader, 1957; Hunt & Otis, 1955). It fails to confirm the results obtained by Scott (1955) who found no differences between ignored animals and those handled for 3 min. per day. In the introduction to the present study it was maintained that the difference in the amount of handling administered by Scott and Weininger could not have accounted for the discrepancy in their results. This was based on the finding that 3 min. of handling was sufficient to reveal differences between handled and nonhandled animals (Ader, 1957). In this preliminary study the open field and Y maze were the test situations. Only the former differentiated between the groups. Although in the same direction, results from the Y maze were not significant. In view of the present observations that only a few of the many tests employed revealed statistically significant results, while those that remained only confirmed the direction of the emotionality among the groups on those tests showing significance, it seems that the amount of manipulation could have accounted for the difference in results.

In Group E_V we have animals exposed to an emotion-provoking situation equal in distribution and duration to the handling procedure, but in this case the emotion-provoking stimuli varied on successive occasions. Group E_D was found to differ significantly from Group E_V only in the greater number of times animals of Group E_V entered concentric Circle 3 of the open field. The only other situations differentiating between the several groups were defecation in the open field and activity in the Y maze. In the former, Groups E_D and E_V did not differ. Both were more emotional than handled animals on the one hand, and less emotional than control animals on the other. In the Y maze also, Groups E_D and E_V did not differ between themselves.

On those tests which did not yield statistically significant results, Groups E_D and E_V performed in much the same manner, as indicated by the very small and nonsignificant difference in the sum of the rankings over the various tests. Both groups tended to be less emotional than control animals; however, only Group E_D with the slightly *lower* rank differed significantly from the controls. Group E_V with the slightly higher ranking did *not* differ from handled animals, while the difference between Groups E_D and H was significant. It appears, then, that although there is a tendency for handled animals to be somewhat less emotional than animals subjected to distributed emotion-provoking stimulation, both such treatments tend to effect a heightened threshold for emotional reactivity relative to unmanipulated animals. Again, while there were few differences demonstrated from specific tests, the over-all tendency is consistent with the hypothesis.

That it was a distribution of the emotion-provoking stimuli, as opposed to emotion-provoking stimuli per se that was effective, is indicated by the finding that animals exposed to massed emotion-provoking stimulation were more emotional than Groups E_D and E_V as well as handled animals. This was true in practically every case. On those tests which differentiated between the groups, animals of Group E_M did not differ from control animals. The same was true of the over-all rankings, although this group did tend in the direction of being more emotional than any of the other groups.

The finding that animals exposed to emotion-provoking stimulation equal in distribution and duration to the handling procedure tend to be less emotional than unmanipulated control animals reinforces the interpretation advanced to account for certain previous results. It will be remembered that Griffiths and Stringer (1952) failed to demonstrate differences between groups of animals subjected to noxious stimulation and controls which received approximately the same treatment except for the noxious stimuli. Also, Stanley and Monkman (1956) failed to demonstrate differences in emotionality between animals given electric shock and control animals placed into the apparatus without shock. It was suggested that the control animals of these studies were comparable to the handled animals of a preliminary study (Ader, 1957) and the present experiment. The slight difference in some cases and the lack of differences in emotionality in others between handled animals and those exposed to the distributed emotion-provoking experience, and the superiority of these groups over controls, are results similar to those obtained by Chevalier and Levine (1955) and Levine et al. (1956). Taken together, these results argue for the inadequacy of the control groups manipulated by Griffiths and Stringer, and Stanley and Monkman, and support, in the main part, the hypothesis that both handled animals and those exposed to emotion-provoking stimulation will display less "emotional" behavior than control animals. The fact that animals subjected to the distributed emotion-provoking stimulation tend to be somewhat more emotional than handled animals is also consistent with the fact that, in Chevalier and Levine's studies, animals given electric shock sometimes differed from no-shock, handled animals. Although in both experiments the experimental groups performed similarly with respect to unmanipulated animals, both experiments also suggest that these experimental groups were somehow different. Whether this difference, should it be a real one, is quantitative or qualitative could not be answered from the data of the present experiment.

The results discussed thus far bear some similarity to those obtained by Griswold and Gray (1956). Working on a physiological level, these investigators hypothesized that sublethal autonomic stimulation would effect a heightened threshold (in terms of mortality rate) to subsequent traumatic stimulation in the Noble-Collip drum. Using electroconvulsive shock as the autonomic stimulant, their hypothesis was confirmed. Pretreatment with ECS reduced the mortality rate due to tumbling trauma in the experimental animals. It had previously been found that sublethal doses of trauma in the drum resulted in the development of resistance to such lethal stimulation (Noble, 1943; Noble & Collip, 1942). That resistance could be built up by previous stimulation which was qualitatively different from the traumatic stimulation to be imposed suggests that the experimental animals do not build up resistance in terms of the transfer of some specific experience.

The question of the influence on emotionality of the repetition of an emotion-provoking stimulus to which the animals may be able to adapt, as opposed to stimulation which does not allow of such adaptation, is answered by reference to Groups E_D and E_V. The latter group, subjected to a variety of noxious stimuli, was not different from those given the same emotion-provoking stimulation on successive occasions. In fact, the difference between these groups is less than that between any others. It appears, then, that whatever part of the heightened reaction threshold to emotion-provoking situations may be described as a transfer effect is a transfer of some general rather than specific adaptive response made to the previously experienced emotion-provoking stimulation.

Of the behavioral measures, only the test involving first emergence from cage revealed an interaction between experience and the time at which it occurred. While nonmanipulated control animals required about the same amount of time to emerge from their cages whether tested early or late in life, all of the experimental groups manipulated early in life were less emotional than their late manipulated counterparts. This finding that the effects of experience vary with the time of manipulation was not confirmed by any of the remaining behavioral situations.

In the case of the second emergence-from-cage test, activity in the open field, entries into concentric Circle 3 of the open field, and latency in the Y maze, animals tested early in life were less emotional than animals tested later in life, regardless of the type of manipulation experienced. With the exception of activity in the Y maze, the remaining measures also showed a tendency for animals tested early in life to be the less emotional.

With regard to the behavioral measures in this experiment, then, the hypothesis that manipulation early in life is particularly crucial in effecting a heightened reaction threshold to emotion-provoking situations was not confirmed. The results indicate that young animals respond to the situations employed with less of what may be described as emotional behavior regardless of the type of manipulation experienced or whether they were manipulated at all.

With regard to the rate of growth of the several groups as measured by the percentage of increase in body weight over the initial weight, no differences were demonstrated. While consistent with Scott's (1955) findings, such results fail to add support to the present hypothesis and fail to confirm the results of other investigators (Bernstein, 1952; McClelland, 1956; Weininger, 1956; Weininger et al., 1954). That animals which were handled and those which were exposed to the distributed emotion-provoking stimulation late in life (Groups H, E_D, and E_V) showed an initial loss in body weight at the beginning of the period of manipulation is in line with the suggestion that these types of stimulation are, initially at least, responded to in the same way. When Group E_M was first manipulated they, too, showed a decrease in body weight. As was pointed out, however, too much emphasis should not be placed upon such data due to the loss in body weight shown by all groups at some time during the period of manipulation.

It is not surprising that animals manipulated early in life did not show this initial weight loss. The rapid growth characteristic of young animals would most probably have covered up such a tendency. Moreover, it is possible that this initial tendency to lose weight in response to the external stimulation is responsible for the observation that when handled animals are heavier than controls, the weight difference first becomes manifest (or becomes greater) late rather than early during the period of manipulation (Greenman & Duhring, 1923; Weininger, 1956).

Weininger (1956) was unable to demonstrate differences between gentled and nongentled animals in the percentage of loss in body weight after 48 hr. of immobilization and food and water deprivation. In the present study, also, the percentage of loss in body weight due to stress did not differentiate between the several groups. It is of interest, however, that the relative position of the groups for both Parts A and B was similar to the rankings for emotionality, with Groups E_M and C showing the greatest loss in body weight.

In a previous study (Herrington & Nelbach, 1942) it was shown that animals subjected to various noxious stimuli ("disturbed" animals) showed a greater body weight than nonmanipulated controls. The greater body weight of "disturbed" animals relative to controls is similar to the relation between gentled and nongentled animals in Weininger's (1956) study. The greater weight of the adrenals of "disturbed" animals relative to controls is not like Weininger's study in that the nongentled animals in his experiment showed the greater adrenal weight. In view of these observations it was suggested that, should greater resistance to stress be displayed by handled animals and those groups exposed to the distributed emotion-provoking stimulation, the physiological mechanisms mediating such resistance in these groups might be different.

From the behavioral tests in the present study it was observed that, although both handled animals and those exposed to the distributed emotion-provoking stimulation tended to be less emotional than controls, these groups differed slightly among themselves. Chevalier and Levine (1955) and Levine et al. (1956) also found that, while differing from unmanipulated controls, their experimental groups did differ somewhat from each other. These observations are consistent with the suggestion made above concerning different mechanisms mediating resistance to stress in handled and "emotionalized" animals.

On a physiological level, however, the results did not, in general, offer reliable evidence for the existence of differences between the several groups, nor did they offer convincing evidence that the mechanisms called into action by the stress were different in the case of any of the groups. Of course, even had differences in resistance been demonstrated to exist, a difference in the mechanisms mediating such resistance might not have been revealed by such a simple measure as the weight of the various organs.

Although the respiration data were, in general, insufficient for complete analysis, the patterns obtained over the 48-hr. period of stress did not even suggest any consistency which might have been associated with a particular group. The initial respiration measure, representing the expected high rate characteristic of an emergency reaction, did not differentiate between the groups, nor did the respiration rate recorded at the end of the stress period. The same was true for the measure of body temperature which was only recorded for animals manipulated late in life. Whether there are, in fact, no differences to be discerned, or whether these observations are due to the great variability shown by these functions (let alone the sparseness of the data) cannot be said with certainty. Although not bearing upon the aims of the present research, the frequently observed tendency for the patterns of change of respiration rate and body temperature to progress in a similar manner over the 48 hr. of stress seems worthy of note.

Because of the observations of Grant (1950), Bartlett, Bohr, and Helmendach (1954), and Bartlett, Bohr, Helmendach, Foster, and Miller (1954) that the initial response of the body temperature to a stress situation is a hypothermic one, and the suggestion that the degree of hypothermia is related to the emotionality of the animal (Bartlett, Bohr, Helmendach, Foster, & Miller, 1954), special attention was given to the first 8 hr. of stress. Contrary to the results cited above, the initial response to the immobilization was a hyperthermic one. Furthermore, the differences in emotionality observed in the behavioral tests did not reveal themselves in terms of the degree of hyperthermia or the time taken before the body temperature returned to the initial level.

These results, however, are in accord with observations on man. Goodell, Graham, and Wolff (1950) found an increase in body temperature to follow upon muscular exercise, and, under a stressful situation, the time required before this temperature returned to the previously established baseline was increased. Such a rise in body temperature under stress has been termed "neurogenic hyperthermia" by Friedman (1950) who describes it as "a phenomenon that spontaneously erupts in patients exhibiting a particular type of autonomic instability primarily characterized by sympathetic over-reactivity to internal or external stress" (p. 440).

Although not directly relevant to the present hypothesis, the data on blood clotting also deserves further discussion. Contrary to what might be expected on the basis of Cannon's (1932) emergency reaction, the time required for the blood to clot was greater in the case of stressed than nonstressed animals. While there are numerous studies which report a shortening of the coagulation time of individuals exposed to emotion-provoking stimulation or stress, there are relatively few studies reporting a lengthening of the clotting time, and there are no studies known to the author which treat of this subject in a systematic manner.

Although they did not discuss it in theoretical terms, Cannon and his co-workers were aware that emotion-provoking stimulation sometimes resulted in an increased clotting time (Cannon & Mendenhall, 1914; Gray & Lunt, 1914). These incidental observations, usually discussed in a footnote, were accounted for in more or less the same way, i.e., "In many experiments the decreased coagulation time rises again not only toward normal as expected, but above. This secondary rise appears due in some cases to reaction in excess, so often remarked in biological processes when opposing factors are at work" (Gray & Lunt, 1914, p. 336), or, the increased clotting time is due "to the use of animals which had been kept for some time in conditions likely to discharge the adrenal glands" (Cannon & Mendenhall, 1914, p. 260).

More recent literature, as reviewed by Selye (1951), does not contribute a solution to the problem of determining the conditions under which there is an increase or decrease in the time required for the blood to clot. The position taken here is that Cannon's (1932) discussion of the increase in clotting time is essentially correct. It is felt, however, that such observations merit a more complete analysis, and the approach recommended is one of determining the changes in blood clotting time over extended periods of emotion-provoking stimulation. Although seemingly obvious, the past history and stimulus employed should be well controlled, precautions which have apparently been ignored (see especially Haden, Schneider, & Underwood, 1948; Scott & Crosby, 1954).

As in the case of HCl secretion in stress situations (Mahl, 1952, 1953), blood-clotting time may be found to vary with the chronicity of the stress and to provide additional support for theoretical analyses and practical therapy based on the distinction between acute and chronic disorders. Another line of research may lie in the personality correlates of alterations in the blood-clotting mechanisms. For example, Macht (1952) has observed that the clotting time of highly nervous blood donors is significantly faster than that of calm individuals, an observation consonant with Cannon's position. Schneider and Zangari (1951) also report decreases in blood-clotting time to a number of *acute* stress situations, but they also found that "when a subject's feeling state could be described as hopeless, sad, depressed, or defeated" (p. 300), the clotting time was significantly lengthened. In this same study they report that while adrenalin injections effect a decrease in clotting time, noradrenaline injections result in a lengthening of the process. The above observations seem somewhat at odds with Funkenstein's (1955) distinction between "anger-in" (depression) with its adrenalin-like reaction, and "anger-out" (rage) with its noradrenalin-like reaction.

The final measures of the present study were of the weights of the various organs. In terms of the degree of resistance to the stress of immobilization and food and water deprivation displayed by the groups given the various types of manipulation, the results offered no support for the suggestion that emotion-provoking stimulation as well as handling would result in a greater resistance to stress than would be evidenced by the nonmanipulation of animals.

As might be expected, the organ weights of stressed and nonstressed animals differed in most cases. There was a decrease in the weight of the heart and liver regardless of the type of manipulation or the time at which it occurred. Because of the involvement of the liver in the blood-clotting mechanism it should be mentioned that the loss in weight of this organ was due, in the main, to the release of stores of glycogen in response to the food and water deprivation accompanying immobilization. Moreover, it has been demonstrated that the prothrombin time is not appreciably affected even when 50% of the liver is damaged (Martin, 1945). It is not likely, therefore, that the clotting time of stressed animals was influenced by the loss in weight of the liver.

In the case of the kidneys and spleen, the difference between stressed and nonstressed animals was a function of whether stress was imposed early or late in life. In both instances, the weight of the organs of animals stressed early in life was less than that of animals stressed late in life. And while the organs of stressed animals weighed less than that of nonstressed animals, this loss was greater in the case of early manipulated animals.

The weight of the adrenal glands did not differ in animals manipulated early or late in life or as a function of the type of manipulation experienced. These results fail to confirm Weininger's (1954, 1956) observations and, on a larger scale, the observations of numerous studies concerning the adrenal glands in the organism's response to stress (see Selye, 1950). The reason for these findings is not apparent. The simplest explanation might be that the stress situation was not sufficiently severe to effect adrenal enlargement. Fortier (1949), however, found a significant decrease in the weight of the adrenals after only 38 hr. of immobilization. This observation, Weininger's results, and the findings of the present study with respect to the loss in weight of other organs would not make plausible the hypothesis that the stress was not sufficiently severe.

One could go to the other extreme and hypothesize that the stress situation was so severe as to cause, or begin to cause, changes in the various organs which indicate complete exhaustion. It has been found, for example, that there is a *decrease* in the weight of the adrenal glands of individuals who were exposed to an intense, continuous stress which ended in death (Uotila & Pekkarinen, 1951). It appears that "in experimental animals and in man, continuous severe stress will eventually exhaust the adaptation energy of the body and lead to a 'stage of exhaustion'" (Selye, 1951, p. 26). The lack of differences in the adrenal weights of stressed and nonstressed animals in the present study may be due to sacrificing the animals just at the time when the hypertrophied adrenal glands were displaying a decline in weight signaling exhaustion. Since no histological or chemical analyses were performed, there is no corroborative evidence for such a suggestion. It is made all the more reasonable, however, in light of the increase in blood-clotting time, which indicates that the animals were no longer responding to an *acute* situation, but one under which the adrenals had become nearly exhausted (Cannon & Mendenhall, 1914).

If one makes the above assumption, then it is not surprising that differences were not demonstrated to exist among the groups. As with an overdose of thiourea (Scott, 1955), all animals succumbed under the severity or chronicity of the situation.

The difficulty with this suggestion is that, to be consistent, one would have to reinterpret Weininger's results and maintain that the heavier weight of the adrenals of nonhandled animals was indicative of sustained resistance rather than greater damage, a conclusion which would appear to be at odds with the remainder of his findings (emotionality, cardiovascular damage, and gastrointestinal erosions). Using only gross measures, the cardiovascular changes noted by Weininger might be characteristic of either the initial response to the stress or a stage of exhaustion (Selye, 1950). The fact that handled animals showed virtually no erosions along the gastrointestinal tract, however, is not characteristic of animals in a state of physical exhaustion and suggests one further argument in an attempt to interpret the present findings in terms of the severity of the stress situation and yet allow for the conclusion that the handled animals in Weininger's study were the more resistant.

In order for the present results and those of Weininger to be consistent it is postulated that the differences in the stress situations employed were sufficiently great so that the animals in the present study were further along in their ultimate collapse under the stress than was the case in Weininger's study. There appear three main reasons why this might be so. In the first place, there are those differences in animals resulting from differing strains and care in different laboratories. Secondly, in Weininger's two experiments involving immobilization, the stress was imposed when the animals were 79 days of age in one, and 59 days of age in the other. In the present study, animals of Part A were immobilized at 87 days of age and animals of Part B at 200 days of age. Finally, although immobilization by means of adhesive tape and gauze was attempted, the wire mesh employed proved considerably more adequate in terms of reducing the amount of struggling. In this situation, the head also was completely immobile, which was not the case in Weininger's study. If, on the basis of these variables, the stress situations in the two studies may be considered sufficiently different, then it is possible that, despite the equal period of stress, greater damage was experienced by the animals in the present experiment.

In theorizing about the results obtained by Weininger, Bovard (1954) has postulated that the differences between handled and nonhandled animals is due to a reduction in handled animals of the activity of the sympathetic nervous system to emotion-provoking stimulation. Similarly, Griswold and Gray (1956), attempting to account for the observations that sublethal stimulation in the Noble-Collip drum effects a resistance to subsequent traumatic stimulation in that situation, have confirmed their hypothesis that a repetition of autonomic activity can result in a gradual alteration in the character of the process so that sympathetic activity is decreased upon subsequent exposure to traumatic stimulation.

Since there is independent research which would lead one to believe that animals react to handling as an emotion-provoking experience (Tryon et al., 1941a, 1941b), and emotion-provoking stimulation of the kind used in this study does elicit autonomic types of activity (Vogt, 1954) one might predict that both handled animals and animals subjected to emotion-provoking stimulation would show greater resistance to stress than unmanipulated controls, and that this resistance is effected by the previously experienced autonomic stimulation.

The failure to find differences in the resistance to stress of the various groups in the present study offers no support for the hypotheses of Bovard and Griswold and Gray. Nevertheless, the results from the behavioral tests are suggestive, since both handled animals and those subjected to the emotion-provoking stimulation tended to be less emotional than nonmanipulated animals. It should not be overlooked, however, that these experimental groups differed somewhat among themselves.

If one assumes that the heightened reaction threshold to the emotion-provoking situations resulted from the previous autonomic stimulation, then it would appear that the degree of autonomic activity effected by handling and emotion-provoking stimulation is different, or that there is more involved here than simply the reactivity of the autonomic nervous system. Of course, one need not make this assumption, but rather, hypothesize that the lesser emotionality or increased resistance displayed by these groups was a result of entirely different mechanisms. Theoretically, these suggestions could be investigated by measuring the autonomic activity of the several groups at the time of manipulation.

SUMMARY AND CONCLUSIONS

This study was designed (*a*) to test the hypothesis that both handling and emotion-provoking stimulation of a nontraumatic nature and distributed throughout early life (as opposed to manipulation at maturity) is capable of raising the threshold for emotional reactivity to subsequent emotion-provoking situations and (*b*) to investigate the effects of handling and various other types of emotion-provoking stimulation on subsequent resistance to stress.

Split-litter groups of animals, representing the following five types of experience, received 30 days of manipulation early (beginning at 23 days) or late (beginning at 136 days) in life.

Group H was handled for approximately 3–4 min. per day in a nonsystematic manner.

Group E_D animals were placed individually into a cylindrical wire cage suspended from an overhanging beam. In this position animal and cage were lifted and allowed to swing freely.

Group E_V was exposed to a variety of emotion-provoking stimuli which differed from day to day. For both Groups E_D and E_V the emotion-provoking stimulation was equal in duration to the handling procedure.

Group E_M was exposed to the same emotion-provoking stimulation as Group E_V; however, in this case they were all presented on the same day.

Group C served as controls, experiencing no manipulation by the experimenter other than the proximity associated with their care and being weighed weekly, as were the experimental groups.

Fifteen days after the last day of manipulation, all animals were tested for emotionality in an emergence-from-cage test, an open-field situation, a Y maze, and a second emergence-from-cage test. After a week without further testing, animals from each group were weighed and then immobilized in wire mesh and remained so for 48 hr. without food or water. Those animals from each group (stress controls) which were not subjected to the immobilization were not manipulated during this time.

During the 48-hr. period of stress, a record of respiration rate was taken. For animals manipulated late in life, body temperature was also recorded. At the end of the stress period all animals were weighed, injected with nembutal, and sacrificed. The heart, liver, kidneys, spleen, and adrenal glands were removed and placed in coded jars of formaldehyde where they remained for approximately one week before being weighed. For late manipulated animals the time required for their blood to clot was also recorded.

Only a few of the many measures obtained from the behavioral tests of emotionality differentiated between the several groups. The relative performance of the several groups in terms of greater or lesser emotionality, however, was significantly constant from all the measures on one test to all those of another, and the direction of these differences were consistent with the results of those measures differentiating between the groups, and with the hypothesis. The over-all emotionality of animals subjected to the massed emotion-provoking stimulation did not differ from that of control animals, but was greater than that for the other experimental groups. The two groups given the distributed emotion-provoking stimulation did not differ. Although there was a tendency for handled animals to be somewhat less emotional than animals subjected to distributed emotion-provoking stimulation, both such treatments tended to effect a heightened threshold for emotional reactivity relative to unmanipulated animals.

The hypothesis that manipulation *early* in life is particularly crucial in effecting a heightened reaction threshold to emotion-provoking situations was not confirmed. The results indicated that young animals respond to the situations employed with less emotionality regardless of the type of manipulation experienced or whether they were manipulated at all.

The rate of growth and loss in body weight under stress did not differentiate between the several groups, offering no support for the hypothesis. There was evidence, however, which suggested that both handled animals and those given emotion-provoking stimulation tend to lose weight as an initial response to the manipulation.

No support for the hypothesis was afforded by any of the physiological measures obtained during or after the stress situation. The lack of physiological differences between the groups was accounted for by postulating that the stress situation was so severe that all animals were in a state of exhaustion at the end of the 48-hr. period.

Incidental to the hypothesis being tested, it was observed that the pattern of change of respiration rate was similar in a given animal to that of body temperature over the 48 hr. of stress. Also, the blood-clotting time of stressed animals was greater than that of nonstressed animals.

REFERENCES

Ader, R. Effects of early experience on emotionality. *Amer. Psychologist,* 1957, **12,** 410.

Anderson, E. E. The interrelationship of drives in the male Albino rat. III. Interrelations among measures of emotional, sexual, and exploratory behavior. *J. genet. Psychol.,* 1938, **53,** 335–352.

Anderson, J. E. Personality organization in children. *Amer. Psychologist,* 1948, **3,** 409–416.

Baron, A., Brookshire, K., & Littman, R. A. Infantile trauma and adult learning in the white rat. Paper read at the meetings of the Western Psychol. Ass., Berkeley, Calif., March, 1956.

Bartlett, R. G., Jr., Bohr, V. C., & Helmendach, R. H. Ability of the rat to adapt to stress of light restraint. *Proc. Soc. exp. Biol., N. Y.,* 1954, **86,** 395–396.

Bartlett, R. G., Jr., Bohr, V. C., Helmendach, R. H., Foster, G. L., & Miller, M. A. Evidence of an emotional factor in hypothermia produced by restraint. *Amer. J. Physiol.,* 1954, **179,** 343–346.

Beach, F. A., & Jaynes, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.,* 1954, **51,** 239–264.

Bernstein, L. A note on Christie's: 'experimental naïveté and experiential naïveté.' *Psychol. Bull.,* 1952, **49,** 38–40.

Bernstein, L. A reply to Scott. *J. abnorm. soc. Psychol.,* 1956, **53,** 141–142.

Bindra, D. What makes rats hoard. *J. comp. physiol. Psychol.,* 1948, **41,** 397–402.

Bovard, E. W., Jr. A theory to account for the effects of early experience on viability of the Albino rat. *Science,* 1954, **120,** 187.

Cannon, W. B. *The wisdom of the body.* New York: Norton, 1932.

Cannon, W. B., & Mendenhall, M. L. Factors affecting the coagulation time of blood. I. The graphic method of recording coagulation used in these experiments. *Amer. J. Physiol.,* 1914, **34,** 225–231.

Chevalier, J. A., & Levine, S. The effects of shock and handling in infancy on adult avoidance learning. *Amer. Psychologist,* 1955, **10,** 432.

Creskoff, A. J., Fitz-Hugh, T., Jr., & Farris, E. J. Hematology of the rat—methods and standards. In E. J. Farris & J. Q. Griffiths (Eds.), *The rat in laboratory investigation.* Philadelphia: Lippincott, 1949.

Doolittle, R. F., & Meade, R. The effect of gentling on some psychological and physiological phenomena. Paper read at the meetings of the Eastern Psychol. Ass., New York, April, 1957.

Duncan, D. B. Multiple range and multiple F tests. *Biometrics,* 1955, **11,** 1–42.

Fortier, C. The effect of atmospheric carbon dioxide on adrenal cortical hyperplasia and associated changes due to stress. *Proc. Soc. exp. Biol., N. Y.,* 1949, **70,** 76–78.

Friedman, M. Hyperthermia as a manifestation of stress. *Proc. Ass. Res. Nerv. & Ment. Dis.,* 1950, **29,** 433-444.

Funkenstein, D. H. The physiology of fear and anger. *Sci. Amer.,* 1955, **192,** 74–80.

Goodell, H., Graham, D. T., & Wolff, H. G. Changes in body heat regulation associated with varying life situations. *Proc. Ass. Res. Nerv. & Ment. Dis.,* 1950, **29,** 418–432.

Grant, R. Emotional hypothermia in rabbits. *Amer. J. Physiol.,* 1950, **160,** 285–290.

Gray, H., & Lunt, L. K. Factors affecting the coagulation of the blood. *Amer. J. Physiol.,* 1914, **34,** 332–351.

Greenman, M. J., & Duhring, F. L. *Breeding and care of the albino rat for research purposes.* Philadelphia: Wistar Inst. Anat. Biol., 1923.

Griffiths, W. J., Jr., & Stringer, W. F. The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.,* 1952, **45,** 301–306.

Griswold, R. L., & Gray, I. Conditioning of rats to tumbling trauma by electro-convulsive shock. Walter Reed Army Inst. Res., *Res. Rep. 182–56,* Nov., 1956.

Haden, R. L., Schneider, R. H., & Underwood, L. C. Abnormal hemorrhage with normal platelet count and normal clotting. *Ann. N. Y. Acad. Sci.,* 1948, **49,** 641–646.

Hall, C. S. Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. comp. Psychol.,* 1934, **18,** 385–403.

Hall, C. S. A comparison of the weights of the adrenal, thyroid and pituitary glands of emotional and non-emotional strains of rats. *Psychol. Bull.,* 1939, **36,** 564.

Hall, C. S. Genetic differences in fatal audiogenic seizures between two inbred strains of house mouse. *J. Heredity,* 1947, **38,** 2–6.

Hall, C. S., & Whiteman, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.,* 1951, **44,** 61–66.

Hammett, F. S. Studies of the thyroid apparatus. I. The stability of the nervous system as a factor in the resistance of the albino rat to the loss of the parathyroid secretion. *Amer. J. Physiol.,* 1921, **56,** 196–204.

Hammett, F. S. Studies of the thyroid apparatus. V. The significance of the comparative mortality rates of parathyroidectomized wild Norway rats and excitable and nonexcitable albino rats. *Endocrinology,* 1922, **6,** 221–229.

Hartman, F. A., & Brownell, K. A. *The adrenal gland.* Philadelphia: Lea & Febiger, 1949.

Hebb, D. O. The effects of early experience on problem solving at maturity. *Amer. Psychologist,* 1947, **2,** 306–307.

HERRINGTON, L. P., & NELBACH, J. H. Relation of gland weights to growth and ageing processes in rats exposed to certain environmental conditions. *Endocrinology,* 1942, **30,** 375–386.

HIGGINSON, G. D. The after-effects of certain emotional situations upon maze learning among white rats. *J. comp. Psychol.,* 1930, **10,** 1–10.

HUMPHREY, G., & MARCUSE, F. New methods of obtaining neurotic behavior in rats. *Amer. J. Psychol.,* 1939, **52,** 616–619.

HUNT, H. F., & OTIS, L. S. Conditioned and unconditioned emotional defecation in the rat. *J. comp. physiol. Psychol.,* 1953, **46,** 378–382.

HUNT, H. F., & OTIS, L. S. Restricted experience and "timidity" in the rat. *Amer. Psychologist,* 1955, **10,** 432.

HUNT, J. McV. The effects of infant feeding frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.,* 1941, **36,** 338-360.

HYMOVITCH, B. The effects of experimental variations on problem solving in the rat. *J. comp. physiol. Psychol.,* 1952, **45,** 313–320.

KARN, H. W., & PORTER, J. M., JR. The effects of certain pretraining procedures upon maze performance and learning. *J. exp. Psychol.,* 1946, **36,** 461–469.

KRUSKAL, W. H. A nonparametric test for the several sample problem. *Ann. math. Statist.,* 1952, **23,** 525–540.

LEVINE, S., CHEVALIER, J. A., & KORCHIN, S. J. The effects of early shock and handling on later avoidance learning. *J. Personality,* 1956, **24,** 475–493.

LOEWE, L., ROSENBLATT, P., & HIRSCH, E. Venous thromboembolic disease. *J. Amer. med. Ass.,* 1946, **130,** 386–393.

MACHT, D. I. Influence of some drugs and of emotions on blood coagulation. *J. Amer. med. Ass.,* 1952, **148,** 265–270.

McCLELLAND, W. J. Differential handling and weight grain in the albino rat. *Canad. J. Psychol.,* 1956, **10,** 19–22.

MACKENZIE, J. B., & MACKENZIE, C. G. Production of pulmonary edema by thiourea in the rat, and its relation to age. *Proc. Soc. exp. Biol., N. Y.,* 1943, **54,** 34–37.

MAHL, G. F. Relationship between acute and chronic fear and the gastric acidity and blood sugar levels in Macaca mulatta monkeys. *Psychosom. Med.,* 1952, **14,** 182–210.

MAHL, G. F. Physiological changes during chronic fear. *Ann. N. Y. Acad. Sci.,* 1953, **56,** 240–249.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.,* 1947, **18,** 50–60.

MARTIN, J. D. Wounds of the liver. *Ann. Surg.,* 1945, **125,** 756–766.

MONTGOMERY, K. C. The effect of hunger and thirst drives upon exploratory behavior. *J. comp. physiol. Psychol.,* 1953, **46,** 315–319.

NOBLE, R. L. The development of resistance by rats and guinea pigs to amounts of trauma usually fatal. *Amer. J. Physiol.,* 1943, **138,** 346–351.

NOBLE, R. L., & COLLIP, J. B. A quantitative method for the production of experimental traumatic shock without hemorrhage in unanesthetized animals. *Quart. J. exp. Physiol.,* 1942, **31,** 187–199.

RUEGAMER, W. R., BERNSTEIN, L., & BENJAMIN, J. D. Growth, food utilization and thyroid activity in the albino rat as a function of extra handling. *Science,* 1954, **120,** 184.

RUEGAMER, W. R., & SILVERMAN, F. R. Influence of gentling on the physiology of the rat. *Proc. Soc. exp. Biol., N. Y.,* 1956, **92,** 170–174.

SCHNEIDER, R. A., & ZANGARI, V. M. Variations in clotting time. *Psychosom. Med.,* 1951, **13,** 289–303.

SCOTT, J. H. Some effects at maturity of gentling, ignoring, or shocking rats during infancy. *J. abnorm. soc. Psychol.,* 1955, **51,** 412–414.

SCOTT, R., JR., & CROSBY, W. H. Changes in the coagulation mechanism following wounding and resuscitation with stored blood: A study of battle casualties in Korea. *Blood,* 1954, **9,** 609–621.

SELYE, H. *Stress.* Montreal: Acta, 1950.

SELYE, H. *First annual report on stress.* Montreal: Acta, 1951.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1946.

STANLEY, W. C., & MONKMAN, J. A. A test for specific and general behavioral effects of infantile stimulation with shock in the mouse. *J. abnorm. soc. Psychol.,* 1956, **53,** 19–22.

STECKLE, L. C., & O'KELLY, L. I. Persistence of response as a function of thirst in terms of early experience with electric shock. *J. comp. physiol. Psychol.,* 1941, **32,** 1–9.

STONE, C. P. Wildness and savageness in rats of different strains. In K. S. Lashley (Ed.), *Studies in the dynamics of behavior.* Chicago: Univer. of Chicago Press, 1932. Pp. 3–55.

TRYON, R. C., TRYON, C. M., & KUZNETS, G. Studies in individual differences in maze ability. IX. Ratings of hiding, avoidance, escape, and vocalization responses. *J. comp. physiol. Psychol.,* 1941, 32, 407–435. (a)

TRYON, R. C., TRYON, C. M., & KUZNETS, G. Studies in individual differences in maze ability. X. Ratings and other measures of initial emotional responses of rats to novel inanimate objects. *J. comp. physiol. Psychol.,* 1941, **32,** 447–473. (b)

UOTILA, U., & PEKKARINEN, A. The relation of human adrenal glands to the pathological changes produced by intense continuous stress ending in death. *Acta Endocrinologica,* 1951, **6,** 23–50.

VOGT, W. The role of the adrenal gland in homeostasis. *Quart. J. exp. Physiol.*, 1954, **39**, 245–252.

WEININGER, O. Mortality of albino rats under stress as a function of early handling. *Canad. J. Psychol.*, 1953, **7**, 111–114.

WEININGER, O. Physiological damage under emotional stress as a function of early experience. *Science*, 1954, **119**, 285–286.

WEININGER, O. The effects of early experience on behavior and growth characteristics. *J. comp. physiol. Psychol.*, 1956, **49**, 1-9.

WEININGER, O., MCCLELLAND, W. J., & ARIMA, R. K. Gentling and weight gain in the albino rat. *Canad. J. Psychol.*, 1954, **8**, 147–151.

WILCOXON, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1941.

WITT, G., & HALL, C. S. The genetics of audiogenic seizures in the house mouse. *J. comp. physiol. Psychol.*, 1949, **42**, 58–63.

WOLF, A. The dynamics of the selective inhibition of specific functions in neurosis: A preliminary report. *Psychosom. Med.*, 1943, **5**, 27–38.

YEAKEL, E. H., & RHOADES, R. P. A comparison of the body and endocrine gland (adrenal, thyroid and pituitary) weights of emotional and nonemotional rats. *Endocrinology*, 1941, **28**, 337–340.

(Accepted for publication February 27, 1958)

Vol. 73, No. 3 Whole No. 473, 1959

Psychological Monographs: General and Applied

QUASI-THERAPEUTIC RELATIONS IN SMALL COLLEGE AND MILITARY GROUPS[1]

FRED E. FIEDLER, EDWIN B. HUTCHINS, AND JOAN S. DODGE[2]

University of Illinois

THE INVESTIGATION here reported deals with the relation of interpersonal pereption to level of adjustment and to change n adjustment.

Previous research has shown that some people characteristically see comparatively ore similarity between themselves and eir associates than do others, and that the ttitude reflected by these interpersonal perptions represents an important variable in erpersonal relations. For example, differences have been found in the perceptions of schizophrenic and normal individuals (Jackson & Carr, 1955). Similarly, one study has shown that reputedly good therapists perceive their patients as more similar to themselves than do reputedly poor therapists (Fiedler, 1951). The present studies relate these phenomena of interpersonal perception to personality adjustment of members of small groups.

Clinicians and personality theorists lean heavily on projective theory and perceptual phenomena in attempting to understand and diagnose personality structure and dynamics. We ask, for example, how effectively various interpersonal attitudes differentiate the well from the poorly adjusted individuals. Since certain of these attitudes appear to be characteristic of relatively good therapists we ask whether these are intrinsically therapeutic attitudes, that is, will we find increased adjustment to occur in any interpersonal situation in which these attitudes play a part? If so, this should permit us to differentiate between a psychologically healthy, "quasi-therapeutic" relation and an interpersonal environment which is detrimental to adjustment.

There is clearly a considerable need for research in this area. As our society has become more complex and the individual's position in it less stable and secure, it has also become more difficult for many to maintain personal adjustment. Although this is due partly to our greater awareness of mental disorders, our society may also be less favorable to the maintenance of mental health. In any case, the demand for mental hospitals and for professionally trained counselors and therapists can no longer be filled, and we must therefore look for methods which prevent or alleviate psychologically maladjustive processes before they need extensive professional intervention. This requires (*a*) that we identify the individual who is unable to cope with his environment and (*b*) that we either place him in an interpersonal situation where he will be able to adjust, or that we help him to modify his interpersonal relations to facilitate his adjustment.

---

[1] This research was conducted under Contract DA-49-007-MD-569 between the University of Illinois and the Human Resources Branch, Research and Development Division, Office of the Surgeon General, and constitutes Interim Technical Report No. 7 of the Project "The Influence of Interpersonal Relations on Effectiveness and Psychological Adjustment of Group Members." Reproduction in whole or in part is authorized for any purposes of the United States Government.

We are deeply indebted to the many who collaborated with us on the collection and treatment of the data, as well as to those whose assistance and support during the various phases of our work made these investigations possible.

Of the project staff, R. E. Jones, Dorothy McBride Kipnis, Betty F. Mannheim, Margaret Reeves Rich, C. F. Wrigley, W. A. Cleven, E. R. Ostrander, Eileen Golb Potter, J. S. Terwilliger, and Audrey Vetter worked with these data at various phases of our research; L. J. Cronbach, N. H. Eisen, D. W. Fiske, and Eleanor Godfrey served as consultants to the project.

We wish to express our thanks to Melvin Manis for permission to use his data, to F. H. Turner and C. S. Sifferd of the University of Illinois for permission to conduct research in the Men's Residence Halls, to L. Dyke and W. M. Gilbert for giving us access to counseling and health center criterion data; to L. E. McDonald and T. R. Vallance, then of Human Research Unit No. 1, Fort Knox, for assistance in testing tank crew members; and to M. S. Carter and his staff as well as to L. J. Davis, A. Kramer, P. Petrilli, Donald Ellis, and C. V. Robbins for permission and cooperation in testing members of the Fifth AAA Regional Command.

We are especially grateful to David McK. Rioch, of the Walter Reed Institute for Research, and C. S. Gersoni for their support, advice, and liaison which enabled us to conduct this work.

[2] Now at the Department of Child Development and Family Relationships, Cornell University.

This issue is of particular importance in organizations such as the armed forces, colleges, and industry. Here a large number of people must frequently work under conditions which are psychologically stressful and disturbing. An understanding of the group factors which enable the individual to adjust or which, on the other hand, are detrimental to his adjustment, is thus of practical as well as of theoretical significance in our society.

Background of the Research

This research extends a number of previous studies by the senior author. The first of these compared the therapeutic relationships established by a group of expert therapists with those of nonexperts from various orientations or schools of psychotherapy. Four psychoanalytic, four nondirective, and two Adlerian therapists participated in the study. Half the members of each group were nationally recognized experts; the others, although they had received all or most of the training required by their "schools," were relative beginners. The results indicated that the experts created therapeutic relationships which were more similar to a generally accepted "ideal relationship" than did the nonexperts. In addition, the experts' relations were more similar to those of other experts than they were to the nonexperts within their own theoretical orientations. The study suggested, therefore, that experts differed from nonexperts primarily in type of therapeutic relationship which they established (Fiedler, 1950a, 1950b).

A second series of earlier studies involved the investigation of therapists' perceptions of their patients. Twenty-two therapists from a large mental hygiene clinic participated. They differed widely in their competence and experience, ranging from highly experienced therapists to relatively inexperienced ones. Each therapist was asked to describe himself and one of his patients by means of *Q*-sort technique. Results showed that the reputedly good therapists tended to see more similarity between themselves and their patients than did their less highly rated colleagues (Fiedler, 1958). A further study by Fiedler and Senior (1952) showed that patients of good therapists tended to idealize[3] their therapists more than did patients of relatively poor ones. These studies raised the question of whether the attitudes reflected by assumed similarity would lead to similar results outside the therapeutic situation.

A subsequent investigation of 26 men living in a fraternity house added further to our understanding of Assumed Similarity (AS) scores. The subjects (*S*s) described themselves, and they predicted the self-descriptions of the person in the group whom they liked best, and the person whom they liked least. As hypothesized, *S*s assumed significantly more similarity to their most preferred person than to the one whom they preferred least.

This relation between AS scores and sociometric preference suggested that AS scores measure along a psychological distance dimension. High Assumed Similarity toward a specific other is presumed to indicate a feeling of warmth, closeness, and acceptance toward him. Low Assumed Similarity, on the other hand, reflects a relatively cold, critical, and rejecting attitude. A number of studies have recently supported this interpretation. Thus, Hausman and his associates[4] conducted an extensive interview and questionnaire study on 172 West Point cadets. They found that cadets who had closer relations with their fathers than with their mothers assumed themselves to be more similar to their fathers, while the reverse was true for men who were closer to their mothers. Steiner and Peters (1958) found persons with high AS to be more conforming and also more concerned with the feelings of their partners. These and similar studies (Davitz, 1955; Fiedler, 1958; Lundy, Katovsky, Cromwell, & Shoemaker, 1955; Precker, 1952) strongly suggested that interpersonal perception scores measure highly important aspects of the interpersonal relationship.

Empirical data and clinical evidence of the importance of acceptance and permissiveness on the part of therapists led to the hypothesis that high AS in social relations is related to good adjustment and that it is conducive to improvement in adjustment. Underlying our approach was the assumption that personality is a product of social interaction and that any set of interpersonal relationships can potentially lead to changes in personality (cf. Witmer & Kotinsky, 1952). Hence, any relationship may be potentially therapeutic or detrimental to adjustment.

---

[3] Idealization was measured in terms of the correspondence between the patient's ideal-self description and his prediction of the therapist on the same questionnaire items.

[4] W. Hausman, B. T. Wiest, and R. Sandison, U. S. Army Hospital, U. S. Military Academy, West Point, N. Y.

In summary, the present study explored certain factors in day-to-day relationships which might be associated with good adjustment and with change in adjustment in order to determine how one individual, or a group of them, can affect another's ability to cope with his environment.

Aims of the Present Research

Our first aim was to discover the relation between interpersonal perception and level of adjustment at a particular time. While it is currently assumed that such a relationship exists, there are few studies which have, in fact, attempted to provide evidence of this relationship or to establish the magnitude of the relation. Our second aim was the investigation of informal quasi-therapeutic relations among members of small face-to-face groups. It was based on the hypothesis that certain aspects of these relations, measured by AS scores, are conducive to change toward better adjustment.

METHOD

Measuring the effect of interpersonal relations on personality adjustment presents a variety of methodological difficulties. Personality changes are readily observable in children and adolescents or in individuals who have suffered brain damage or severe psychological deprivations. Changes occurring in normally functioning individuals—that is, in nonclinical groups—are at best very subtle phenomena which occur slowly and almost imperceptibly. Even deliberate attempts to change personality, such as in psychotherapy on the one hand and "brainwashing" on the other, require as long as one, two, or three years. In contrast, we attempted to identify modifications in personality which were not deliberately induced, and which took place in a time period of weeks rather than years.

Furthermore, our primary interest centered on effects which were general rather than those which might be specific to one group or one situation. For these reasons, our tests had to be limited to determining whether or not a relationship exists, and whether or not changes did in fact take place, rather than estimating the magnitude and permanence of the changes.

The Samples

The over-all plan of the investigation required the comparison of two samples of student groups and two samples of military units. During the first year of our work, we utilized data which had been collected on university students housed in 14 dormitory suites[5] ("The dormitory groups"). In addition, we collected data on members of 40 army tank crews ("The tank crews"). During the second year, the original procedures for data collection and criterion development were repeated, as far as possible, on two further samples of *S*s. These were students housed in 13 separate temporary barracks ("The barracks groups") and on members of 53 anti-aircraft artillery units ("The anti-aircraft artillery crews").

To permit the measurement of change over time, as well as adjustment at a particular period, *S*s were tested twice. In all but the "old" anti-aircraft artillery units the first testing occurred shortly after the group members had become acquainted with one another. The *S*s were tested a second time after the relationships had become more clearly structured (see Fig. 1). To increase the readability of this discussion, the two studies on college students are described first, the two on military units, later.

The dormitory groups. These were 14 groups of 107 male college students living in a new university dormitory. Most *S*s had entered the college at the beginning of the semester in which the data were collected. About half the students were freshmen, the remainder being transfer and incoming graduate students. Most *S*s lived in rooms accommodating four men. Two rooms were generally joined

---

[5] These data were collected by Melvin Manis for his dissertation (Manis, 1955). We are indebted to him for the raw data of this study.

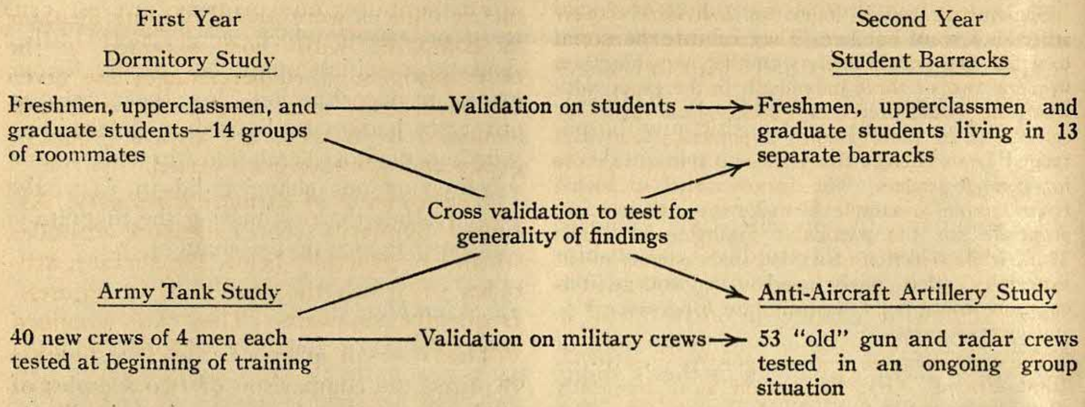

First Year
Dormitory Study
Freshmen, upperclassmen, and graduate students—14 groups of roommates
Validation on students
Second Year
Student Barracks
Freshmen, upperclassmen and graduate students living in 13 separate barracks
Cross validation to test for generality of findings
Army Tank Study
40 new crews of 4 men each tested at beginning of training
Validation on military crews
Anti-Aircraft Artillery Study
53 "old" gun and radar crews tested in an ongoing group situation

FIG. 1. Schematic representation of the over-all research plan.

to form a section which constituted the unit of our study. The *S*s were assigned rooms in order of their application for housing and very few had known each other before the start of the semester. All were tested during the fifth week of the semester and again during the eleventh week. Our analyses were based on 87 *S*s for whom complete data were available. Due to illness or withdrawal from school the remaining *S*s were unavailable at the time of the second testing session.

The barracks groups. Data similar to the above were collected on a sample of 172 upper-class and freshmen students living in 13 temporary barracks. Each barracks, which was divided into four-man suites, held a maximum of 16 men. Analyses were restricted to 101 freshmen for whom complete data could be obtained. These *Ss* were tested during the third and twelfth weeks of the semester.

The tank crews. This part of the investigation was conducted on members of a tank battalion during the crew training phase. Each crew consisted of five enlisted men who were new to each other and to the particular type of tank training. Average age of the 160 men initially tested was 22 years, 49 were married and three were divorced. Average education was 9.7 years; mean Army General Classification Test score was 90. This group was, therefore, considerably more heterogeneous than the college student groups. Attrition due to transfers, discharges, overseas levies, and other causes reduced the retest sample size from 160 to 95 men.

The anti-aircraft artillery crews. In contrast to the tank crew members who were mostly recruits in training, the last group of *S*s came from "old" operational anti-aircraft artillery units. These men were no longer in their initial training period, but were performing their regular duties within the Air Defense Command. Units varied in size from six to fifteen men. As with tank crews, the loss of data, due to furloughs, sick calls, overseas levies, or personnel turnover, was considerable. In the 12-week period between testings, the original sample shrank from 483 men to 200 for whom complete data were available. Our analyses were based on this group of 200 *S*s. Here again there was considerable heterogeneity among *S*s. The average age of the men, all of whom were enlisted, was 23 years, 36 were married; average education was 11.2 years; and the mean Army General Classification Test score was 101.

It is obvious from the attrition rates in our samples that it would have been impracticable to extend the time between testing sessions. Even in the much more stable college populations, an attempt to carry the investigation beyond one semester's duration was completely frustrated by the number of students who had left school, moved to new quarters, or who, for other reasons, were no longer available for testing.

Measurement of Interpersonal Relations

On the basis of previous work on psychotherapeutic relations, we considered two kinds of indices. These were (*a*) sociometric preference ratings and (*b*) measures of psychological distance, viz., AS scores.

Sociometric indices. Sociometric rating forms were designed for two purposes. First, we wanted to identify for each *S* the "significant others" in the group. Specifically, we wanted to determine the persons with whom *S* might have a quasi-therapeutic relationship. Secondly, we wished to determine each *S*'s sociometric status. This latter purpose in collecting sociometric ratings was to provide some indication of *S*'s personal adjustment and will be discussed in the next section of this report.

Sociometric scores index an individual's overt attitudes toward another. They indicate the extent to which one person prefers another by asking him to name two or three individuals in the group with whom he would prefer to associate in a variety of activities. In these studies hypothetical activities ranged from choices for week-end guests to those for combat leaders. The forms varied somewhat from sample to sample to make the questions appropriate for the particular group under study. Thus, while selection of a good study partner might materially affect one's progress in college, this question would not be appropriate for members of anti-aircraft crews.

The questionnaires were designed to identify *S*'s "Best Friend" (F), the person most frequently chosen in response to the various questions, and a quasi-therapeutic figure, here called the "Confidant" (C). The latter was defined by the question, "If you had some personal problems you wanted to talk over with someone in your group, with whom would you most likely discuss them?" It should be noted here that the confidant was simply the person whom *S* chose in response to this question. The confidant need not, in fact, have acted in this capacity. Where we asked *S*s who had come to them with personal problems, we found that some persons named as confidants were not consulted, and that some who had actually been asked for their advice did not recognize this fact (see Appendix A and Appendix B for sociometric questionnaires).

There was, of course, considerable overlap between confidant and friend choices. This was inherent in our definition of the "best friend" as the person who was sociometrically most preferred. This individual was quite likely to be chosen as a confidant as well. As Table 1 shows, the overlap is between one-half and two-thirds of the choices, except in the case of the tank crew study. Here, the groups were so new and the personnel changes so great that sociometric choices were probably quite unstable as compared with those obtained in the other samples. Two percentages are given for each study since confidants or friends identified by *S* were not always from *S*'s group and *N*'s therefore varied.

TABLE 1

OVERLAP BETWEEN SUBJECTS' FRIEND AND CONFIDANT CHOICES IN THE VARIOUS STUDIES

Study	Friends Who Were Also Confidants (%)	Confidants Who Were Also Friends (%)
Dormitory	60	65
Barracks	54	65
Tank	37	24
Anti-aircraft artillery	41	67

Measurement of assumed similarity. Assumed Similarity scores measure attitudes which are generally below the surface, attitudes of which the individual is, himself, frequently unaware. They are obtained when we ask a person to describe himself on a set of items descriptive of personal attributes and then using the same test items to describe others. Previous studies have shown that the item content plays a relatively slight role in the measurement of AS (Fiedler, 1958).

The scale sheet used in the dormitory study contained 24 items which were based on Cattell's (1950) factor analysis of Allport and Odbert's list of adjectives used by English-speaking people in describing themselves and others. The 24-item tests were subsequently modified and a 20-item scale sheet was utilized in the three later studies.

The form of the items followed that of Osgood's Semantic Differential (Osgood, Suci, & Tannenbaum, 1957). While other test formats have been used for obtaining AS scores, Osgood's appears to be most efficient. Items consisted of six-point scales bound by pairs of adjectives and their antonyms (see Appendix B). A seven-point scale was used in the dormitory study (see Appendix A). Examples of these items are as follows:

Friendly _√_ : ___ : ___ : ___ : ___ : ___ : Unfriendly
Cooperative ___ : _√_ : ___ : ___ : ___ : ___ : Uncooperative
Stable ___ : _√_ : ___ : ___ : ___ : ___ : Unstable

The *S* described himself (or another person) by checking the space in each scale which most adequately represented his judgment of that person. Approximately two to three minutes were required for the completion of a 20-item scale sheet.

Items were identical for each description *S* made. Following appropriate instructions he described: (*a*) himself; (*b*) himself as he ideally would like to be (his ideal self); (*c*) three or more close work companions; and (*d*) his confidant.

Computation of AS scores. Assumed Similarity scores require the comparison of two scale sheets to determine the profile similarity of the two descriptions.

In the preceding illustration of the scales, assume that the check marks might represent an *S*'s self-descriptions. Thus he would have marked the three items at Points 1, 2, and 2. Assume that he described his coworker by marking the corresponding items at Points 5, 4, and 6 as presented below.

Item	Description of self	Description of other	Difference between self and other	Difference squared
Friendly	1	5	4	16
Cooperative	2	4	2	4
Stable	2	6	4	16

$$\Sigma D^2 = 36$$

$$D = \sqrt{\Sigma D^2} = 6$$

D, the statistical measure of profile similarity used here, is then obtained by extracting the square root of the sum of the squared item differences (Cronbach & Gleser, 1953). In this case, the squared differences would be 16 on the first and third items, and four on the second, yielding a sum or D^2 of 36. D thus equals six. Note that the amount of similarity assumed by the rater is inversely proportional to the size of the D score. The more a person sees another as resembling himself, the more the two descriptions will coincide and the smaller will be the D score.

Assumed Similarity scores can be computed between any pair of scale sheets an *S* fills out. In the present studies we concentrated primarily on those scores which might be related to quasi-therapeutic interactions. The most important of these, as Mannheim (1957) has shown, reflects the interaction between the individual and his group. They indicate how close the individual feels to the members of his group, and to what extent he is accepted by them. Similarly, Asch's (1952) studies on conformity behavior have shown that the opinion of the group is an extremely important factor in changing the individual's attitudes and thinking. In the present study we therefore used two average AS scores to and by the group as the main predictors. The first of these, *S*'s average AS to the members in his group ($\overline{\mathrm{ASg}}$), indexed the degree to which *S* felt close to, and identified with, those with whom he lived and worked. The second, the average AS of group members toward the subject ($\overline{\mathrm{ASs}}$), indicated the extent to which he was seen as one of the group, i.e., the extent to which he was accepted by its members. These scores, being averages, were also quite reliable.

We had planned to work with four additional scores which were expected to predict adjustment and changes in adjustment. These were the individual's acceptance and idealization of his confidant, and his confidant's acceptance and idealization of him. (Idealization was operationally defined as the D score between the *S*'s ideal self and his description of another person. It indicates how closely *S* sees the other as resembling *S*'s own ideal.)

The relationship between the subject and his confidant was the nearest analogue to the patient-therapist relationship with which our data provided us. We had therefore expected that this set of indices would yield the best predictions. However, these scores, as will be shown later in Table 3, were highly correlated with $\overline{\mathrm{ASg}}$ and $\overline{\mathrm{ASs}}$; they were not treated as main predictors in these studies.

We also wished to determine whether it was the confidant relationship which had special significance, or whether any close relationship with another "significant" person would serve the same function. To determine this, indices analogous to confidant relationships were computed for *S* and his best friend. The indices used are summarized below:

Symbol	Interpretation of AS Score
$\overline{\mathrm{ASg}}$	Average similarity which *S* assumes between himself and other members of his group with whom he has close working or friendship relations.
$\overline{\mathrm{ASs}}$	Group members' average AS to *S*, indicates the extent to which *S* is accepted by others in his group.
ASc	*S*'s AS to his confidant. *S*'s psychological distance from the person he chooses as potential confidant.
ASIc	*S*'s idealization of his confidant, i.e., the similarity *S* assumes between his ideal self and his confidant.

Symbol	Interpretation of AS Score
ASf	*S*'s AS to his "best friend," i.e., the similarity *S* assumes between his most preferred group member and himself. This score was obtained to compare friendship with confidant relations.
ASIf	*S*'s AS between his ideal self and his "best friend." Score obtained to compare confidant and social relations.
C's ASs	Confidant's AS toward *S*, indicative of his acceptance of *S*.
C's ASIs	Confidant's idealization of *S*.
F's ASs	Friend's AS toward *S*, his acceptance of *S*.
F's ASIs	Friend's idealization of *S*.

Reliability and Intercorrelations of AS Scores

Table 2 presents the internal consistencies of the various scores which were used in this investigation. The coefficients were obtained by Guttman's split-half method. Table 3 presents the intercorrelations obtained in the four studies. To simplify the table we have presented median correlations. Thus, ASc and ASIc which listed as .64 in Table 3 is the median of .66, .30, .62, and .72 for the dormitory, barracks, tank, and anti-aircraft artillery studies, respectively. This matrix is further summarized in Table 4 which also indicates the presence of two distinct clusters of AS scores: (*a*) scores obtained from the same *S* about others and (*b*) scores obtained from others about a particular *S*. It should be noted that the scores within each cluster are mathematically related to one another and the intercorrelations are, therefore, artifactually high. For example, the average scores, $\overline{\mathrm{ASg}}$ and $\overline{\mathrm{ASs}}$ are based in part on ASc and ASf on the one hand, and C's ASs and F's ASs on the other. Furthermore, the ASc and ASIc scores are interrelated since both are based in part on *S*'s description of his confidant. This made it advisable that we restrict our main analyses to the unconfounded and independent scores, $\overline{\mathrm{ASg}}$ and $\overline{\mathrm{ASs}}$. The remaining scores are presented for illustrative purposes.[6]

TABLE 2

RELIABILITIES FOR TEN PREDICTOR SCORES IN THE ANTI-AIRCRAFT ARTILLERY SAMPLE

(Time I Scores)

Symbol	*r*	*N*
S's AS to Others		
ASc	.79	200
ASIc	.79	200
ASf	.81	145
ASIf	.83	145
$\overline{\mathrm{ASg}}$	.86	200
Other's AS to *S*		
C's ASs	.78	89
C's ASIs	.85	89
F's ASs	.84	145
F's ASIs	.89	145
$\overline{\mathrm{ASs}}$	.84[a]	168

Note.—*r*'s are corrected by the Spearman-Brown prophecy formula.

[a] Estimated from C's and F's ASs scores and the split-half reliability of group members' AS scores.

CRITERIA OF ADJUSTMENT

This research dealt with individuals who were forced to live as members of groups. The *S*s were chosen because their group membership provided information which is difficult to obtain from the private clinic patient or from the individual who works and lives alone.

The ultimate criterion of adjustment has here been conceptualized in terms of the individual's ability to participate fully in his group. This ability has been called the individual's "staying power." Conversely, maladjustment has been defined as the extent to which the individual cannot, or does not, participate effectively in his interpersonal situation.

If a person is maladjusted by our definition of the term, he may withdraw by physically removing himself from the group, e.g., by being absent without leave, by evidencing psychoneurotic reactions which

---

[6] As Cronbach has pointed out, AS scores are complex measures which might be more meaningfully interpreted by analyzing them into their component variances. While this method was not adopted in the present research, it seems very likely that a more refined approach of this nature would lead to better prediction and perhaps also to more precise interpretation of the results.

TABLE 3

MEDIAN PREDICTOR INTERCORRELATIONS FROM ALL FOUR STUDIES

	ASc	ASIc	ASf	ASIf	$\overline{\text{ASg}}$	C's ASs	C's ASIs	F's ASs	F's ASIs	$\overline{\text{ASs}}$
ASc		.64	.80	.48	.69	.12	.04	.05	.09[a]	.18
ASIc			.42	.59	.52	.19	.12	.06	.05[a]	.13
ASf				.59	.74	.03	.10	.09	.11[a]	.21
ASIf					.43	.06	.06	.10	.08[a]	.15
$\overline{\text{ASg}}$						.03	.08	.05	.08[a]	.15
C's ASs							.68	.74	.38[a]	.46
C's ASIs								.48	.80[a]	.50
F's ASs									.70[a]	.52
F's ASIs										.48[a]
$\overline{\text{ASs}}$										

[a] Based on 3 *r*'s only.

limit his participation in group activities, or by manifesting the kinds of character disorders which make him personally obnoxious to other group members who then exclude him from full participation.[7]

Three classes of criteria reflect this notion of staying power. They include (*a*) subjective indices of adjustment, (*b*) indices based on group acceptance, and (*c*) objective or behavioral indices of personal effectiveness. The specific measures used in these studies are described below.

Subjective Indices of Adjustment

The extent to which a person feels well, the extent to which he feels able to cope successfully with his environment, is a most important aspect of adjustment. No matter how "sick" a person may be in the eyes of others, he will generally consult a therapist only when he, himself, feels the need for professional help. Although many an advanced schizophrenic has protested to his physician in the locked ward that he never felt better in his life, self reports by reasonably well-functioning individuals are clearly useful measures of their ability to get along in life. Two subjective indices of adjustment, self-esteem and self-satisfaction, were derived from the rating scales described earlier. Several others, based on more or less disguised adjustment inventories, were also obtained. These indices are discussed below.

TABLE 4

SUMMARY OF MEDIAN PREDICTOR INTERCORRELATIONS INDICATING TWO DISTINCT CLUSTERS OF *S*'S AS TO OTHERS AND OTHER'S AS TO *S*

Study	*S*'s AS to Others	Others AS to *S*	*S*'s AS to Others by Other's AS to *S*
Dormitory	.60	.49	.20
Barracks	.36	.49	.12
Tank	.65	.71	−.01
Anti-aircraft artillery	.65	.62	.10

Self-esteem (SE). The score, SE, was the sum of *S*'s self-ratings on each scale when the items were scored on a good–bad continuum. It indicated the degree to which *S* esteemed or valued himself. To calculate SE, the favorable pole for each scale was first determined. In the example given earlier, this appears at the left of each scale. We assumed that it is better to be friendly than unfriendly, stable than unstable, and cooperative than uncooperative.[8]

The favorable end of the continuum was arbitrarily scored lowest. Thus in the above example, a person marking the friendly–unfriendly scale at the point at which the check mark appears would receive a score of "1" on that item; he would receive a "2" on the second and third items. Since SE is the sum of the item scores, it would here equal "5." The lower the score, the higher the self-esteem.

---

[7] Obviously, a well adjusted individual might not like a particular group, and he might try to transfer out of it, or he may be promoted. We assume here that such a person could have stayed in the group, if he had desired to remain.

[8] Three judges agreed unanimously as to which was the positive pole for all 20 scale items. This a priori determination of the "good" pole was empirically confirmed in the anti-aircraft artillery study. The *S*'s responses on the ideal self-description were tallied and a frequency distribution made for each item. All but one of these distributions were highly skewed with the modes of each falling on the scale value previously assigned a "1."

Several studies have indicated a positive relation between the favorability of the self percept (or other self-report measures) and various criteria of adjustment (Brownfain, 1952; Murray, 1938). High self-esteem indicates that the individual values himself highly and sees himself as being adequate in terms of the cultural values. He describes himself as being characterized by the traits our culture considers desirable.

Self-satisfaction. A second index of the favorability of the self percept was the Self-Satisfaction score. This measured the correspondence between the descriptions of self and ideal self and thus indexed the degree to which an individual perceived himself as having the characteristics *he* most admired. A similar score has been used extensively by Rogers (1951) and by Rogers and Dymond (1954) as well as in earlier studies (Fiedler & Senior, 1952). These indicated that self-satisfaction is related to various adjustment criteria. In our own work, we found it significantly correlated with self-esteem. (The two scores are not mathematically independent, since both were derived in part from the same data, viz., *S*'s self-description.) The magnitudes of the correlations between the two sets of scores were not sufficiently high (median $r = .50$), however, to justify retaining only one.

Taylor Manifest Anxiety scores. In all but the dormitory study, a modified form of the Taylor Scale of Manifest Anxiety (Taylor, 1953) was administered. Rosenthal's (1955) findings support the clinical opinion that this scale measures overt signs of anxiety, and other studies have demonstrated that it has adequate reliability and validity (Cattell, 1950). After preliminary tests, 48 of the original items were chosen for inclusion in the present studies.

General Army Adjustment Index. This index, developed by Havron, Fay, and Goodacre (1951), is based on the assumption that an individual who is motivated to remain and succeed in the army is better adjusted to the military situation than is a person who wishes to withdraw. The scale taps such attitudes as one's willingness to go into combat, satisfaction with army training, feeling that the army assignment is worthwhile, etc. It was administered only to members of anti-aircraft units.

Group Acceptance Criteria

The second major aspect of adjustment we considered was the individual's ability to interact with others in a manner which was acceptable to the members of his group. Our data yielded two indices of this ability.

Sociometric status. To some extent, this operational definition of group acceptance tends to become confused with conformance to group norms and "other-directedness," although many studies have shown that sociometric status can be used as an adjustment measure. Typical are the findings of French (1951), and of Lemann and Solomon (1952). (See also Lindzey & Borgatta, 1954.) These results indicate a positive relation between sociometric choice and job performance as well as other criteria of personal adjustment.

The usual method for measuring this variable was followed. The *S*s were asked to indicate those in their group whom they preferred as roommates, combat leaders, etc. Choices given to an individual by the members of his group were summed and, to compensate for differences in group size, adjusted for the total number of possible choices. The higher the score, the more *S* was considered to be accepted by his group members.

Mean esteem by others ($\overline{EO}$). This score was obtained from the rating scale descriptions which others in the group made of *S*. The method for calculating this score was identical with that used in computing the self-esteem index, except that the descriptions by others rather than *S*'s self-description were used. Since an *S*'s final score was the average of all descriptions made of him, the score in effect represented the group's estimate of the individual's adjustment.

Because of the size of some of the groups in our samples, it was not always feasible to have every person describe every other member of the group. In the barracks and anti-aircraft artillery samples, for example, as many as 15 men constituted a group. Thus, we asked every member to describe only the four men with whom he had closest work contacts. Some biasing undoubtedly occurred as a result of this partial sampling procedure. However, this effect was probably not very pronounced inasmuch as most men generally do not have much latitude in choosing their work partners.

Objective Indices of Personal Effectiveness

These criteria reflect the individual's success in attempting to adjust to his life situation. Thus, one essential aspect of adjustment for the college student is his ability to remain in school. Similarly, the adjustment of a man in a military unit must be judged in part by whether or not he can remain in his outfit. If he is beset by mysterious ailments, or if he gets court-martialled and must spend a year in the guard house, it would be difficult to consider him well adjusted to army life.

Useful criteria of this nature are usually specific to the situation under consideration, e.g., school grades are meaningful only in a school situation, court-martial records only in the military services. Objective criteria applicable to all four samples could, therefore, not be obtained. The measures which we were able to obtain in the various samples are discussed below.

Health center visits. Data on the frequency and nature of the individual's visits to the health center were collected on both college student groups. University regulations required all students to undergo a physical examination at the time of their entrance to the university. Their medical history was recorded at that time, as were any specific disorders then present. Throughout the semester, each visit made to the health center was added to the student's record. The notation included the nature of the complaint, the diagnosis, treatment, and any pertinent comments the physician cared to make.

An unusually high frequency of visits was taken as one index of poor adjustment. With the assistance of the medical staff we attempted to differentiate *S*s with symptoms of a psychosomatic nature (headaches, back pains, etc.) from those who had purely somatic difficulties (fractures, etc.). The presence of psychosomatic complaints was taken as a second index.

An analogous index utilizing sick-call visits was obtained from the tank study. (We were unsuccessful in obtaining adequate sick-call data from individuals in the anti-aircraft artillery sample.) For obvious reasons, this measure did not have the same meaning as it did with university students. There is a great deal of secondary gain for a soldier in reporting for sick call which enables him to avoid onerous details. This is, of course, not generally true in the case of students who have to take time out from classes, study periods, or extracurricular activities in order to visit the health center.

Counseling bureau visits. The university's student counseling bureau provided additional information regarding the psychological adjustment of *S*s in the dormitory and barracks samples. The majority of students visited the counseling bureau at least once to learn the results of their performance on the freshman guidance examination. At that time, a counselor interviewed them and rated them with regard to the intensity of personal, educational, and vocational problems. On the advice of the counseling bureau staff, only the presence of personal problems was considered to indicate poor adjustment on this criterion.

While the rated intensity of the difficulty yielded one adjustment score, the number of visits provided another. Students who returned several times for therapy were classified as maladjusted on this criterion. Generally, students who used the counseling bureau for more than the initial visit had also been rated by the counselor as manifesting severe personal problems. Thus, all subjects who consulted the counseling bureau more than once because of personal problems were classified as functioning below average, regardless of the nature of the difficulty. It seems reasonable to assume that students who consistently felt the need for such help in solving their problems were not as well adjusted as those who were able to manage their own affairs with a minimum of professional aid.

College grades. One criterion of personal effectiveness in college was the student's grade point average (GPA) for the semester under study. For better or for worse, a college student's adjustment is measured largely by how well he performs in his work at school. No matter how happy he may be, no matter how delightful his relationships might be with his roommates, his adjustment to the college situation as a whole must be considered unsatisfactory if he fails or gets kicked out of school. This is not to say that the student must necessarily graduate with honors to be considered adjusted, but a person who has scholastic ability, as shown by his high school performance and college aptitude test scores, yet fails to perform well, is not adjusting to the demands of his environment.

Disciplinary ratings. Adequate individual proficiency scores were not available in the military units. Except for sick calls in the tank study, ratings on disciplinary offenses, obtained in the anti-aircraft artillery study provided the only objective indication of the individual's ability to adjust to the army.

Reliability of Criterion Measures

Table 5 presents split-half reliability coefficients for all adjustment indices susceptible to this statistical treatment. (Adequate reliability estimates of indices, such as sick

TABLE 5

SPLIT-HALF RELIABILITY COEFFICIENTS OF PERSONALITY ADJUSTMENT INDICES

Index	Sample	*N*	r^a
Self-satisfaction	Barracks	169	.76
Self-esteem	Anti-aircraft artillery	415	.93
Taylor manifest anxiety	Anti-aircraft artillery	200	.91
General army adjustment	Anti-aircraft artillery	200	.73
Mean esteem by others	Dormitory	86	.70
Sociometric status	Dormitory	99	.94

[a] Corrected for length.

calls or disciplinary ratings, could not be obtained from the data at our disposal.) As relatively little variation was found from sample to sample, only one estimate is presented.

As the data show, self- and group-acceptance indices had sufficiently high internal consistency to serve as criterion measures for this study.

Intercorrelations Among Criterion Indices

Inasmuch as the problem of the interrelations among adjustment criteria has been dealt with more fully in a previous paper (Fiedler, Dodge, Jones, & Hutchins, 1957), detailed analyses are not presented here. Table 6 shows the intercorrelations among the criteria of adjustment that were used. With the exception of artifactual rela-

TABLE 6

INTERRELATIONSHIPS BETWEEN ADJUSTMENT INDICES OBTAINED IN TWO OR MORE STUDIES

(Time II Tables)

Index	Dormitory	Barracks	Tank	Anti-Aircraft Artillery
Self-satisfaction and:				
Self-esteem[a]	.60	.55	.36	.44
Sociometric status	.09	.09	.00	.10
Mean esteem by others	−.02	.15	−.14	−.09
Taylor anxiety[b]	—	.34	.37	.30
Health center	.00	.19	.10	—
Grade point average	.17	.30	—	—
Counseling bureau	.15	.16	—	—
Self-esteem and:				
Sociometric status	.20	.10	.04	.09
Mean esteem by others	.22	.09	−.20	.02
Taylor anxiety[b]	—	.25	.43	.34
Health center	.22	−.06	−.09	—
Grade point average	−.14	.28	—	—
Counseling bureau	.08	.22	—	—
Sociometric status and:				
Mean esteem by others[c]	.38	.42	.34	−.17
Taylor anxiety	—	−.09	.09	.04
Health center	.06	−.04	.02	—
Grade point average	.05	.31	—	—
Counseling bureau	.10	.00	—	—
Mean esteem by others and:				
Taylor anxiety	—	−.02	−.13	−.07
Health center	.15	.08	−.09	—
Grade point average	.03	.14	—	—
Counseling bureau	.23	.06	—	—
Taylor anxiety and:				
Health center	—	.16	.07	—
Health center and:				
Grade point average	−.10	−.06	—	—
Counseling bureau	.06	.02	—	—
Grade point average and:				
Counseling bureau	.03	.27	—	—

[a] Spuriously high relation based on self-descriptions.

[b] Both variables are self-descriptions.

[c] Both variables based on descriptions by same others.

tions, the intercorrelations are quite low. On the basis of these data we concluded that adjustment could not be measured by means of a composite or global index. Rather, the results suggested that the measures indexed separate aspects of adjustment. Thus, our analyses dealt separately with each criterion score.

Table 7 presents the intercorrelations among change measures. Here again the data suggest that changes in personality take place piecemeal. Apparently we cannot, as yet, readily encompass "personality adjustment" by one composite measure. Nor, as the data further suggest, can we talk about personality adjustment as if it were a unitary trait or attribute of a person.

To facilitate interpretation of the data, the basic analyses are restricted to five "main" criterion variables. These variables were chosen: because an adequate estimate of their reliability could be obtained; because they were utilized in at least three of the four studies, thus providing some indication of the generality of the results; and because they could be obtained twice in each study, thus yielding change-in-adjustment scores.

The criterion scores which met these conditions were: (*a*) Self-Esteem, (*b*) Self-Satisfaction, (*c*) Taylor Anxiety Scores, (*d*) Sociometric Status, and (*e*) Mean Esteem by Others. Analyses related to criterion measures not meeting these requirements are reported as subsidiary findings.

RESULTS

This research program dealt with two main questions. The first of these considered whether people who are adjusted perceive others differently than do those who are less well adjusted, and whether they are, in turn, perceived differently. The second question concerned *changes* in adjustment. It asked whether an individual's perception of others or others' perceptions of him were related to subsequent modifications in his adjustment.

Relation of Assumed Similarity to Level of Adjustment: General Criteria

The nature of the relations between psychological distance and personal adjustment were investigated by correlating AS scores with various adjustment indices. Since we

TABLE 7

RELATIONSHIPS BETWEEN INDICES OF CHANGE IN ADJUSTMENT OBTAINED IN TWO OR MORE STUDIES

Index	Dormitory	Barracks	Tank	Anti-Aircraft Artillery
Self-satisfaction and:				
Self-esteem[a]	.46	.43	.18	.38
Sociometric status	−.11	−.10	.00	.08
Mean esteem by others	.15	.03	.09	.04
Taylor anxiety[b]	—	.26	.24	.08
Self-esteem and:				
Sociometric status	.08	−.14	.01	.04
Mean esteem by others	.00	.23	−.13	−.02
Taylor anxiety[b]	—	.23	.01	.09
Sociometric status and:				
Mean esteem by others[c]	.24	.09	.16	−.09
Taylor anxiety	—	.26	.05	−.03
Mean esteem by others and:				
Taylor anxiety	—	.04	.11	−.03

[a] Spuriously high relationship—both variables self-descriptions.

[b] Both variables based on self-descriptions.

[c] Both variables based on ratings by same others.

were primarily interested in effects which are general over different situations, we dealt simultaneously with the relations obtained in all four studies by following a procedure for averaging and weighting each of the correlations by the size of the sample. This procedure yielded an index of the homogeneity of the various correlations over the samples (Edwards, 1950, pp. 133–141). Only where this homogeneity exists—that is, where χ^2 is *not* significant—is one justified in pooling the correlations over several samples. A significant χ^2 indicates significant differences between samples.

As will be recalled, the most general predictor variables were $\overline{\text{AS}}$g (*S*'s average AS to all persons whom he described in the group) and $\overline{\text{AS}}$s (the average amount of similarity assumed toward *S* by his fellow group members). The former is interpreted as indicating *S*'s general acceptance of his group members, the latter the degree to which he is accepted by them.

Table 8 presents the pooled correlations between the two main predictors, and the five main criteria. As can be seen, four of the seven correlations were statistically significant, with three of them (Taylor Manifest Anxiety and $\overline{\text{AS}}$g, Sociometric Status and $\overline{\text{AS}}$g, and Sociometric Status and $\overline{\text{AS}}$s) reaching the .01 level. All were in the expected direction. The first hypothesis was thus confirmed. It should be pointed out that Taylor Anxiety scores and $\overline{\text{AS}}$g are psychologically related since both are dependent to some extent on the individual's self-perception. A person who perceived himself favorably on the self-description would also tend to describe himself favorably on the Taylor Anxiety scale. (The median correlation between favorability on self-description and Taylor Anxiety scores over the four studies was .34.) The self-description enters also into each of the scores reflecting assumed similarity to others. Hence, the findings, while significant, must be interpreted with some caution. (Parenthetically, we might point out that this same limitation holds whenever we intercorrelate any "objective" personality tests which are based on self-reports.)

It has been shown in previous studies that we assume similarity to people we like and esteem. A person who was liked by his peers and who was sociometrically chosen by them was also perceived as similar. The results are, therefore, in line with expectations based on previous work.

Of particular interest are the correlations between $\overline{\text{AS}}$g and Mean Esteem by Others, and $\overline{\text{AS}}$s and Taylor Manifest Anxiety scores. While these correlations are relatively low, they do indicate that *S*'s adjustment is related to his attitudes toward others, and others' attitudes toward him.

We next considered the relations between AS to, and by, *specific* members in *S*'s group. Table 9 presents the correlations

TABLE 8

AVERAGE CORRELATIONS (*r*) BETWEEN GENERALIZED AS SCORES $\overline{\text{AS}}$G AND $\overline{\text{AS}}$S AND THE MAIN INDICES OF ADJUSTMENT

(Time I)

Index	*N*	$\overline{\text{AS}}$g	*N*	$\overline{\text{AS}}$s
Self-esteem	—	—[a]	501	.07
Self-satisfaction	—	—[a]	501	.08
Taylor manifest anxiety	425	.37**	414	.11*
Sociometric status	512	.04	501	.22**
Mean esteem by others	428	.15**	—	—[a]

Note.—Unless otherwise indicated, the χ^2s for homogeneity of correlations over groups were not significant and have been deleted.

[a] These correlations would be spurious and were therefore omitted.

* Significant at the .05 level.

** Significant at the .01 level.

TABLE 9

Average Correlations (r) Between S's Time I AS to Confidants and Friends and Scores on the General Adjustment Indices

Index	N	ASc	ASIc	N	ASf	ASIf
Taylor manifest anxiety	282	.43**	.32**	231	.39**	.27**
Sociometric status	309	−.05	−.02	258	−.04	−.01
Mean esteem by others[a]	247	—	.12	218	—	.11

[a] χ^2 = 7.21 (significant at .05) and 5.25 for ASc and ASf, respectively, indicating that correlations are not homogeneous and cannot be averaged.

** Significant at the .01 level.

between S's AS scores to his confidant and friend, and the general adjustment indices. The only significant relations which can be meaningfully interpreted are those with Taylor Anxiety, all of which reached the .01 level. While not independent of one another, these relations suggest that individuals who idealize and assume similarity to those who are important to them are less likely to exhibit overt signs of anxiety.

Table 10 presents the correlations between criteria and AS toward the subject by specific others. Two of the eight relations involving confidant's AS scores were significant, as was one involving S's best friends. These were the relations between confidant's or friend's perceptions of S and S's sociometric status. Both of these relations were somewhat higher than $\overline{\text{AS}}$s and sociometric status. They suggest that the person who is well liked or highly chosen is more idealized by his confidant and/or friend than is the less chosen individual. Since the confidant's and friend's responses comprise part of the individual's sociometric status score, the finding might mean that the popular individual is able to choose friends and confidants from among those who think highly of him, while the less popular person will not even be idealized by those whom he considers his friends or confidants.

Level of Adjustment: Analyses Using Specific Criteria

The major emphasis has thus far been on adjustment measures which were obtained on three or more of the studies. There were, however, a number of measures which, although specific to only one or two studies, provided some valuable information. These data are presented below.

Grade point average. The relations between students' college grades and AS were analyzed in both college student samples. Since tests are not given until after at least the sixth or eighth week of school, the AS scores are, in effect, predictors of perform-

TABLE 10

Average Correlations (r) Between Confidant's and Friend's Time I AS to S and S's Scores on General Adjustment Indices

Index	N	C's AS	N	C's ASIs	N	F's AS	N	F's ASIs
Self-esteem	246	.01	245	.01	319	.02	236	.03
Self-satisfaction	246	.02	245	.04	319	.04	236	.08
Taylor manifest anxiety	162	.06	161	.01	235	−.01	236	−.06
Sociometric status	246	.14*	245	.30**	319	.07	236	.31*

* Significant at the .05 level.

** Significant at the .01 level.

ance. Table 11 presents the correlations between AS scores and GPA. As can be seen, the only significant relation in the dormitory study was that between GPA and subject's tendency to assume similarity to his confidant. Although this did not reach significance in the barracks groups, it closely approached it. Pooled, these correlations were significant at the .05 level. The main predictor, $\overline{\text{ASg}}$, was significantly related to the criterion in the barracks study as was the tendency to idealize the best friend. Again these relations were not significant in the dormitory study, although the combined correlations were significant.

Obviously, grade point averages contain many sources of variance. We were, however, able to control for the student's native ability. The adequacy of his previous training, the competence and grading systems of instructors, and various other factors, all of which affect student performance, were uncontrolled. Since all of these contribute to error variance, the obtained correlations are probably somewhat lower than the true correlations. Thus, the data suggest that persons who assume similarity to others are more effective in their school work than are those who assume little similarity. Inasmuch as GPA is not known to the student until fairly late in the semester, these correlations are predictive of adjustment as measured by this criterion.

Counseling bureau visits and student health center visits. In both college student samples, the number and nature of visits to the counseling bureau and student health center constituted adjustment criteria. Although there were scattered significant correlations between AS scores and these indices, no consistent relations were found.

Sick calls. Records indicating the number of times the individual went on sick call and the nature of his symptoms could be obtained on tank crew members. The men were housed in a fairly small geographical area, they were seen by their own medical personnel, and the policy regarding sick call was fairly uniform from unit to unit.

Table 12 presents the correlations between sick call visits and AS scores in the tank crew sample. As can be seen, the individual's acceptance of his fellow group members in general and his confidant in particular were significantly correlated with the frequency of his visits. Similarly, a person who was idealized by his friend tended to go on sick call less frequently than a person who was not idealized. All correlations were positive with exception of

TABLE 11

AVERAGE CORRELATIONS (*r*) BETWEEN GPA AND AS SCORES IN THE DORMITORY AND BARRACKS STUDIES

AS Scores	Dormitory		Barracks		Average	
	N	*r*	*N*	*r*	*N*	*r*
$\overline{\text{ASg}}$	87	.15	95	.22*	179	.19*
$\overline{\text{ASs}}$	87	.00	85	.08	179	.04
ASc	87	.22*	95	.16	179	.19*
ASIc	87	.17	95	.17	179	.17*
ASf	87	.12	73	.22	157	.17*
ASIf	87	.00	73	.32*	157	.15
C's ASs	87	.11	39	.09	119	.10
C's ASIs	86	.13	39	−.04	118	.08
F's ASs	87	.12	48	.24	123	.16
F's ASIs	—	—	48	−.07	—	—

* Significant at the .05 level.

TABLE 12

CORRELATIONS (*r*) BETWEEN SICK CALLS (TIME I) AND AS SCORES IN THE TANK STUDY

AS Scores	*N*	*r*
$\overline{\text{ASg}}$	90	.27**
ASc	95	.20*
ASIc	95	.03
ASf	59	.21
ASIf	59	.03
$\overline{\text{ASs}}$	93	−.06
C's ASs	22	.38
C's ASIs	21	.37
F's ASs	24	.36
F's ASIs	25	.41*

* Significant at the .05 level.
** Significant at the .01 level.

$\overline{\text{ASs}}$ and sick call visits. In brief, the soldier who had closer relations with others appears somewhat less likely to go on sick call or to exhibit psychophysiological symptomatology than one whose relations were more tenuous.

General Army Adjustment Index. This index is basically a measure of personal morale and satisfaction with army life. Questions concern the possibility of reenlisting, having one's son become an army career man, feeling that one is performing an important job, etc. These questionnaires were complete only for the anti-aircraft artillery crew members in our study.

Table 13 indicates that all AS-to-others scores were significantly correlated with this index. We concluded, therefore, that the perception of others as similar or dissimilar is related to adjustment to army life. The more the individual perceived others to be similar to himself, the better his adjustment on this criterion.

Disciplinary ratings. Next we obtained disciplinary ratings from each crew leader for every member of his crew. These ratings reflected the amount of "trouble" which a particular man gave his sergeant. A person who is a constant disciplinary problem is obviously not adjusting to the demands of military life.

The correlations between AS scores and these ratings are again presented in Table 13. Being perceived as similar by others ($\overline{\text{ASs}}$) was significantly related to good adjustment on this criterion. The relationship with the confidant is of particular interest. Idealization *of and by* the confidant were associated with ability to adjust to military discipline.

Annoyance scores. Finally, in the tank crews we obtained information as to the

TABLE 13

CORRELATIONS (*r*) BETWEEN GENERAL ARMY ADJUSTMENT, DISCIPLINARY RATINGS (TIME I), AND AS SCORES IN THE ANTI-AIRCRAFT ARTILLERY STUDY

AS Scores	General Army Adjustment		Disciplinary Ratings	
	N	*r*	*N*	*r*
$\overline{\text{ASg}}$	200	.24**	200	.01
$\overline{\text{ASs}}$	200	−.04	200	.19**
ASc	200	.22**	200	.12
ASIc	200	.34**	200	.16*
ASf	200	.17*	200	.04
ASIf	200	.28**	200	.00
C's ASs	89	−.13	89	.16
C's ASIs	89	−.06	89	.30**
F's ASs	145	−.17*	145	.13
F's ASIs	145	−.13	145	.15

* Significant at the .05 level.
** Significant at the .01 level.

number and kinds of things which annoyed the *S*s. Only one correlation reached significance, namely, *S*'s annoyance score and the confidant's AS toward *S*. Since this is the only significant relation in ten, this result is obviously not conclusive.

To summarize, our first hypothesis stated that the interpersonal perceptions of well adjusted individuals differ from the perceptions of relatively poorly adjusted persons. Likewise, we hypothesized that well adjusted individuals are perceived differently by their fellow group members than are the less well adjusted individuals. These hypotheses were generally supported by our data.

The Relation of Assumed Similarity to Change in Adjustment

The second major question concerned the relation between interpersonal attitudes, as reflected by AS scores, and changes in adjustment. We hypothesized that individuals who are perceived as more similar by others in their group improve in adjustment, while those perceived as relatively dissimilar experience a decrement. Likewise, we expected that those who saw others as more similar would adjust more readily than individuals who perceived others to be dissimilar.

A number of methodological problems had to be considered in testing this hypothesis. Change or "growth" scores present a variety of difficulties, particularly in the personality adjustment area. First of all, our scales may not have psychologically equivalent scale intervals. Thus, a shift from a self-satisfaction score of 2 to 4 might have a different psychological meaning than a similar shift from 10 to 12. In other words, a shift on the part of a very dissatisfied individual might mean a considerable change in his life, while a shift of the same magnitude on the part of a very self-satisfied individual might mean relatively little.

More importantly, given measures which are to any extent unreliable, one must control for "regression to the mean" effects. Thus, a person who is extremely well adjusted at a particular time has less chance of showing improvement at a later date than one who is poorly adjusted. Conversely, a person who scores very low at a first testing session will, purely on the basis of probability, tend to improve his score rather than decrease it on further testing (cf. Lord, 1956).

To control for this statistically spurious effect, *S*s were matched on the basis of their Time I criterion scores. All change scores were dichotomized into those above and below the median. The *S*s having the same score on the first testing were then rank ordered on the basis of their change in adjustment on a particular criterion.[9] For each Time I score, *S*s were matched by pairing the *S* who had improved the most since Time I with the *S* who had deteriorated the most. Similarly, the second most improved person was matched with the second most deteriorated. Pairs were formed as long as one score in the pair came from above the median of the total sample and the other score came from below. We thus had pairs of improved (change scores from above the median) and deteriorated *S*s (change scores from below the median) who had been matched on the basis of the adjustment scores at the initial testing session. This procedure is schematized in Fig. 2. The method had one major drawback. Since we eliminated from our sample all those who could not be matched, persons with extremely high and extremely low scores at Time I were not included in this analysis.

Matched pairs analysis of variance tests[10] indicated whether the AS scores obtained from *S* (as well as from his fellow group members) predicted which of the two

---

[9] The score indicating change in adjustment was computed by subtracting the Time II score from the Time I score. Thus, the change could be in a positive or a negative direction.

[10] Analogous to a matched *t*, the matched pairs analysis of variance test was the more convenient technique to use since it also allowed a check on the significance of the between-rows variance, i.e., the relation between Time I criterion scores and the AS measures. *F*-test results from all four samples were then combined by the χ^2 method.

Subject	Time I Score	Time II Score	Change Score	
Jones	2	5	−3	
Smith	2	4	−2	
Brown	2	3	−1	Matched
Johnson	2	1	+1	
Kennedy	3	6	−3	
Roberts	3	5	−2	
Jenkins	3	4	−1	Matched
Caldwell	3	2	+1	
Davidson	3	1	+1	

Improved Subject	Matched With	Deteriorated Subject
Johnson	⟷	Jones
Davidson	⟷	Kennedy
Caldwell	⟷	Roberts

FIG. 2 Matching procedure for analysis of change in adjustment. All names are fictitious.

paired *S*s would improve and which would become less well adjusted.

Table 14 shows the relations between change in adjustment and the main predictors, $\overline{\text{ASg}}$ and $\overline{\text{ASs}}$. Three of the seven relations were statistically significant, two are beyond the .01 level of confidence. Just as in the patient-therapist relations, high AS toward the individual was conducive to his improving in adjustment. In addition, high AS *by* the individual to other members of his group aided his adjustment—a fact not found in the study on patient-therapist relations (Fiedler & Senior, 1952). This latter very likely reflects a situational difference. A patient who perceived his therapist as very similar to himself would be seeing him as poorly adjusted. Obviously, a patient would not stay with a therapist whom he perceived in this way. The present studies dealt with nonclinical populations where a person would not necessarily be seeing himself as poorly adjusted.

TABLE 14

F TESTS SHOWING THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN TIME I AVERAGE AS SCORES FOR *S*S IMPROVED AND DETERIORATED IN ADJUSTMENT[a]

Index	$\overline{\text{ASg}}$ *F*	$\overline{\text{ASs}}$ *F*
Self-esteem	—	.07
Self-satisfaction	—	3.14**
Taylor anxiety scores	2.07*	2.33**
Sociometric status	−.05	.97
Mean esteem by others	−.06	—

[a] Relations in this table and in the succeeding tables were obtained by Matched Pairs Analysis described in the body of this paper.

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

Table 15 presents the relations between change in personality adjustment and AS to friend and confidant. Of these, only ASc and Taylor Anxiety scores reached the .05 level. Since the Taylor Anxiety score correlates also with $\overline{\text{ASg}}$, this result is probably not due to chance.

None of the relationships between personality adjustment and friend's and confidant's AS to the subject reached significance.

To summarize, the data have shown that AS scores are predictive of some aspects of change in adjustment as well as correlates of adjustment level at a particular time. The findings suggest that the attitudes we have been dealing with are of consider-

TABLE 15

COMPARISON OF RELATIONS BETWEEN AS SCORES TO CONFIDANT AND FRIEND AND CHANGE IN ADJUSTMENT

Index	ASc	ASIc	ASf	ASIf
Taylor anxiety scores	2.04*	1.00	1.27	−.47
Sociometric status	−.84	−.32	−1.49	.66
Mean esteem by others	.18	.12	.83	1.40

* Significant at the .05 level of confidence.

able importance in interpersonal relations. While the observed changes in adjustment were small in absolute terms, the fact that they emerged despite very short intervals of time indicates that these interpersonal attitudes must exercise a fairly strong influence on adjustment. The fact that the results were obtained in studies using four different samples, and in two and probably three quite disparate situations, indicates that we were dealing with fairly general effects.

Discussion

This investigation studied (*a*) the relationship between interpersonal perception and level of personal adjustment, and (*b*) the hypothesis that attitudes reflected by high AS scores are conducive to favorable changes in adjustment.

Interpersonal perceptions and the individual's level of adjustment. The results bearing on the first hypothesis were in the expected direction and generally significant. The study showed that individuals having relatively close interpersonal relations with others in the group tended to be better adjusted on the various criterion measures than were more distant individuals. But contrary to prevailing theory as well as reported research (e.g., Bruner & Tagiuri, 1954; Willoughby, 1934), the correlations obtained were fairly low. A number of factors must be considered in interpreting these findings, however. The low magnitude of the correlations might be the result of inefficient measurement of interpersonal attitudes; it might be due to poor criteria of adjustment; or these correlations might reflect the true level of the relationships.

There seems to be little reason for suspecting the efficiency of AS scores. A number of studies have shown them to be very reliable indicators of relevant dimensions in interpersonal relations (Fiedler, 1958; Fiedler & Senior, 1952), and the present investigations show them sensitive to relatively small changes in personal adjustment. On the other hand, we do know from previous work (Fiedler et al., 1957) that the criteria of adjustment are independent. Thus, each index measures only one aspect of personal adjustment and is unrelated to other indices. Moreover, we were dealing with a nonclinical population from which the seriously disturbed individuals had already been eliminated. In light of this, it seems clear that the use of these measures as diagnostic indicators of adjustment is not warranted. Whether further refinement of these measures and the development of better criterion measures will yield more promising results, or whether these correlations reflect the true relationship between these interpersonal perception scores and personal adjustment is, of course, a question for further research.

Interpersonal perception and change in adjustment. The most important contribution made by this study was the identification of quasi-therapeutic interpersonal relationships. We found that *S*s with high AS to others in their group ($\overline{AS}$g) tended to become less anxious, and that those who were perceived as more similar ($\overline{AS}$s) became not only less anxious, but more satisfied with themselves as well.

These findings are important for two reasons. First, they point to a new approach to preventive psychotherapy and to mental health programs in large military and industrial organizations where men must work under conditions which are often stressful and psychologically damaging (cf. Fiedler, 1949). Secondly, these findings suggest a new approach to understanding the psychotherapeutic process. They suggest that we look at therapeutic relations within the context of group interactions rather than from the vantage point of the therapist working intimately and intensively with his patient.

In light of the shortages of professional personnel in the mental health field, ways of supplementing the work of the psychotherapist in office practice are clearly needed. The present approach provides one potential answer by pointing up the possibility of alleviating personal maladjustment within the structure of the individual's work group. Although a great deal of work remains to be done, in broad outline the findings suggest some steps which might lead to the construction of a healthy interpersonal environment for the marginally adjusted individual, or for persons who are in a particularly stressful situation. They indicate that the individual's personal adjustment can be influenced when he is placed in a group whose members accept him (high $\overline{AS}$s), and whom he, in turn, accepts (high $\overline{AS}$g).

In this connection it should be noted that certain aspects of personality adjustment changed considerably over very short periods of time, and under conditions which could hardly be regarded as favorable to therapy. This suggests that adjustment, at least as we measured it, is more labile than has hitherto been supposed. If adjustment can change during an individual's short-lived membership in one group, changes are also likely to occur when he enters another group. Adjustment would then vary with the prevailing group attitudes.

While we do not, as yet, have any definitive way of favorably modifying these group attitudes, our research has shown that it is possible to influence some of the relevant quasi-therapeutic factors. First, AS is a fairly generalized personal attribute. An individual assuming a high degree of similarity to one person will tend to assume much similarity to others with whom he has comparable relationships. To some extent, therefore, we can select the group into which a particular individual should be placed.

Secondly, our own as well as other studies has shown that acceptance or rejection is, to a large extent, related to purely ecological variables. For example, we found that college students assigned to the same room were more likely to choose each other as friends and confidants than to choose persons in neighboring rooms. Similarly, as French (1951), has reported, the individual who, almost by random methods, is assigned to leadership functions, is more likely to attain high sociometric status.

We are suggesting here that it is possible to institute a systematic attempt to reduce the anxiety or to improve the work efficiency of group members by judiciously placing them with others who are basically accepting. That this is not unrealistic has been shown by Bushard (1957) who recently reported that members of small military units can be enlisted to assist in helping one of their maladjusted fellow group members through his crises. That such within-group relationships can exert a powerful influence on the individual is further seen in two previously published case studies of a successful and an unsuccessful quasi-therapeutic relationship (Eisen, 1957). In summary, then, our findings have indicated that steps can be taken to create a therapeutic environment for individuals even while they are members of small work groups. Further work in this area will undoubtedly reveal still better methods for utilizing the group's potential for serving as a quasi-therapeutic agent.

The second major contribution of this research lies in its implications for our understanding of the process of personality change. The problem of how personality changes is a particularly crucial issue in the areas of psychotherapy and personality development. Yet, while the psychotherapist is in an ideal position intensively to observe the changes which take place in one single patient, the validity of generalizations from such data about personality change in general is limited for a number of important methodological reasons.

First, the subjects in any research on psychotherapy come from a highly selected population. This is especially true when the *S*s are patients of therapists in office practice. A major portion of these come from the upper middle class which tends to see psychological difficulties as one of the natural hazards of modern life. Their friends and family generally accept, if they do not encourage, the individual's seeking treatment. In addition, patients in this social class tend to be highly verbal and introspective in contrast to individuals from the lower socioeconomic levels. These latter, according to many clinicians, are more inclined to act out their difficulties. In contrast, the samples in our studies cover a wide spectrum of individual differences. Our *S*s ranged from predominantly upper middle-class college students to relatively unselected men in military organizations. The latter were, of course, drawn from all socioeconomic levels and spanned a wide range of intelligence scores.[11]

Secondly, it is very difficult to compare a large number of cases treated by different therapists. Therapists also generally have very strong loyalties to their own theoretical orientations which makes it difficult for them to interpret data which may be at variance with their points of view. This problem becomes particularly important when attempts are made to explain the dynamics of the therapeutic process. Therapists differ widely in accounting for the patient's movement in therapy, e.g., by interpretation of the transference relationship, or of the life style, or by bringing into the relationship accepting, permissive attitudes. Also, the therapist is enjoined by some to play the role of the stern father, by others to act contrary to how the patient expects people to behave, and by still another group to be a sounding board against which the patient may try out his feelings and ideas in a protected environment.

Our investigations suggest that these specialized attitudes are not necessary for the individual's progress toward better adjustment outside the therapist's office. The quasi-therapeutic relationship, as far as we can determine from our research, is characterized by the similarity which people see between themselves and others, i.e., by the extent to which people within a small group accept each other. This formulation comes perhaps closer to Rogers' ideas than to those of other theorists in the field.

The main question, of course, asks why these perceptions of people should affect personal adjustments. But before discussing this point it should be noted that the changes observed in this study occurred primarily on self-report measures. These are, of course, also the objective criteria used in most research on psychotherapy. Generally speaking, the therapist must rely on the patient's statements that progress is, or is not, being made. Even when the patient is asked to return for a follow-up interview, or a series of psychological tests, we are forced to rely on his introspective accounts of himself and his reactions to others. This is the case whether these reports are in the form of interview responses or whether they are elicited in disguised form by the Rorschach or TAT (e.g., Rogers, 1951; Rogers & Dymond, 1954). Reliance on this type of criterion alone is, of course, unavoidable in most clinical studies. Even if sufficient funds and professional personnel were available to conduct field investigations of patient behavior, few individuals would submit to having a social worker or a therapist interview their employers, coworkers, or relatives to assess their improvement.

---

[11] Psychologists have frequently been accused of basing their science on the white rat and the college freshman. While we did not work with white rats, we did study college freshmen. The fact that very few chi squares in the correlational analyses were significant indicates that college and military groups behaved quite similarly insofar as their reactions to group attitudes were concerned.

While subjective indices of adjustment are thus expedient, it is obvious that they introduce biases into the definition of adjustment and therapeutic success. It is, therefore, of interest that the findings of Rogers (1952), as well as the results of our own research, tend to show shifts primarily on self-report measures rather than on ratings or evaluations made by others. This may mean that the individual does not change very much as far as his observable behavior is concerned, and that the critical changes are those which reflect his subjective feelings of well-being. It is, of course, possible that the time spans between testing sessions in our studies were too short to provide adequate information on the *S*s' observable behavior. Changes may have occurred but escaped the awareness of the, after all, untrained raters. However, in the absence of empirical evidence to the contrary, we must tentatively conclude that interpersonal attitudes affect primarily the individual's subjective feelings about himself. This does not imply that psychotherapy is, for this reason, less valuable. The patient seeks therapeutic help because he feels anxious and beset by doubts about himself—not because he is disliked by others, but because he feels uncomfortable about it. It may require a considerably longer time to change others' attitudes about oneself than to change one's own attitudes about self and others.

While quasi-therapeutic and formal therapeutic relations are probably not interchangeable, they are analogous. With this in mind, we can now attempt to answer the question of why high AS is related to improvement in these aspects of adjustment. In the first place, seeing someone as similar has been shown to mean seeing him in a favorable light (Fiedler, Warrington, & Blaisdell, 1952). People perceived as similar are rated as falling nearer to the favorable poles on the various items in our questionnaires. Hence, our findings suggest that a person comes to perceive himself more favorably when he is perceived as similar, i.e., in a favorable manner, by his group (high $\overline{\text{AS}}$s).

In essence, this finding is consistent with Festinger's (1954) theory of social comparison processes. Festinger hypothesized that "there exists, in the human organism, a drive to evaluate his opinions and his abilities" (p. 117). He postulates further that, "to the extent that objective, nonsocial means are not available, people evaluate their opinions and abilities by comparison respectively, with the opinions and abilities of others" (p. 118). Inasmuch as self-regarding attitudes are opinions of oneself, the theory may be applied directly to our data. Given a situation in which the individual finds that every member of his group regards him in a favorable light, he will then tend to change his opinion to conform to that of the group. Asch (1952) has shown, in fact, that individuals will tend to conform to group opinion even when dealing with the evaluation of a physical stimulus, such as the length of a line. That is, many *S*s will rate an actually longer line as "shorter" when all other members of the group have previously rated the line as shorter. When we deal with a highly ambiguous stimulus, such as a self concept, group opinion is much more likely to affect the individual's judgments about himself.

In a patient-therapist relationship, the patient sees the therapist as well adjusted, while the good therapist sees the patient as quite similar to himself (Fiedler & Senior, 1952). Therapy may then be seen in terms of the patient's gradual acceptance of the therapist's valuation of him. Thus, through interaction with the therapist the patient becomes more secure and more satisfied with himself. Steiner and Peters (1958) found that *S*s assuming a high degree of similarity tended to exhibit greater conformity behavior. The individual who assumes a high degree of similarity should thus also be the one who more easily accepts others' attitudes regarding his own worth. Since two or more persons' opinions should have more weight than the opinion of just one, we might further hypothesize that a patient will also improve more rapidly if he deals with more than one therapist. While no systematic studies on this point have been reported, it has been noted by a number of clinicians whose patients had transferred from other therapists. This would further suggest that group therapy might be managed so as to capitalize on these effects.

Finally, our findings may shed some light on the phenomenon of patients or potential patients who improve without benefit of psychotherapy. If our data may be interpreted freely, we may well have a clue here to at least one cause of these "spontaneous remissions," namely, that the individual was able to get into a group where he could form a relationship in which he was accepted, and where he could perceive himself to be similar to others whom he considered to be well adjusted (cf. Fiedler, 1949).

In summary, it is our belief that research directed toward discovery of the day-to-day interpersonal relations which contribute to better adjustment, or which lead to spontaneous remission, holds the key to better community mental health programs and preventive psychotherapy. This investigation has shown that quasi-therapeutic relationships can be identified, and it has suggested some of the elements which contribute to change in adjustment. It is our hope that this investigation will lead to more intensive research in the future.

SUMMARY AND CONCLUSIONS

The present investigation dealt with the hypothesis that a relationship exists between interpersonal perception and the level of personal adjustment, as well as changes in adjustment.

Four groups were utilized. Two of these consisted of college students living in university dormitories and temporary housing facilities. Two further groups consisted of members of army tank crews and antiaircraft artillery units.

Assumed Similarity (AS) scores were obtained from all *S*s at two sessions, spaced from six to twelve weeks apart. The AS scores are obtained when an *S* describes himself, as well as a number of others in his unit, on identical personality questionnaires. Descriptions are then compared by means of the D Statistic to determine the degree of similarity or difference which *S* sees between himself and others.

A number of criterion indices were used for purposes of measuring personal adjustment. Previous work had shown that these adjustment indices were essentially uncorrelated, and the present investigations, therefore, treated each separately.

Results of this study indicated that AS scores, especially those indicative of the individual's feeling of closeness to the group, and the group's acceptance of the subject, are related to adjustment criteria. We found statistically significant, albeit very small, relations between AS scores and level of adjustment. Statistically significant differences in changes in adjustment were also obtained between groups of *S*s paired on the basis of their level of adjustment at the first testing session.

These findings were discussed in terms of previous research in this area, as well as in terms of their implications for theories of personality and psychotherapy, and the potentiality of developing programs of preventive psychotherapy and mental health.

REFERENCES

ASCH, S. Effects of group pressure upon the modification and distortion of judgments. In E. S. Swanson, T. M. Newcomb, and E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1952.

BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.,* 1952, **47,** 597–606.

BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 2. *Special fields and applications.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 634–654.

BUSHARD, B. L. The mental hygiene consultation service in the army. Paper read at the symposium on preventive and social psychiatry, Walter Reed Army Institute of Research, Washington, D. C., April 15-17, 1957.

CATTELL, R. B. *Personality: A systematic, theoretical, and factual study.* New York: McGraw-Hill, 1950.

CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.,* 1953, **50,** 456–473.

DAVITZ, J. Social perception and sociometric choice of children. *J. abnorm. soc. Psychol.,* 1955, **50,** 173–176.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

EISEN, N. H. Quasi-therapeutic relationships: A clinical case study. Urbana, Ill.: Group Effectiveness Res. Lab., Univer. of Illinois, *Interim Tech. Rep. No. 4,* 1957.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.,* 1954, **7,** 117–140.

FIEDLER, F. E. An experimental approach to preventive psychotherapy. *J. abnorm. soc. Psychol.,* 1949, **44,** 389–393.

FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective, and Adlerian therapy. *J. consult. Psychol.,* 1950, **14,** 436–445. (a)

FIEDLER, F. E. An investigation into the concept of the ideal therapeutic relationship. *J. consult. Psychol.,* 1950, **14,** 239–249. (b)

FIEDLER, F. E. A method of objective quantification of certain countertransference attitudes. *J. clin. Psychol.,* 1951, **7,** 101–107.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana, Ill.: Univer. of Illinois Press, 1958.

FIEDLER, F. E., DODGE, JOAN S., JONES, R. E., & HUTCHINS, E. B. The measurement of personality change and adjustment in nonclinical populations. Urbana, Ill.: Group Effectiveness Res. Lab., Univer. of Illinois, *Interim Tech. Rep. No. 5,* 1957.

FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.,* 1952, **47,** 446–453.

FIEDLER, F. E., WARRINGTON, W. G., & BLAISDELL, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.,* 1952, **47,** 790–796.

FRENCH, R. L. Sociometric status and individual adjustment among naval recruits. *J. abnorm. soc. Psychol.,* 1951, **46,** 64–72.

HAVRON, M. D., FAY, F. J., & GOODACRE, D. M., III. *Research on the effectiveness of small military units.* PRS Rep. No. 885. Inst. for Res. on Hum. Relat., 1951.

JACKSON, W., & CARR., A. Empathic ability in normals and schizophrenics. *J. abnorm. soc. Psychol.,* 1955, **51,** 79–82.

LEMANN, T. B., & SOLOMAN, R. L. Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry,* 1952, **15,** 7–90.

LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. *Theory and method.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 405–448.

LORD, F. M. The measurement of growth. *Educ. psychol. Measmt.,* 1956, **16,** 424–437.

LUNDY, R. M., KATKOVSKY, W., CROMWELL, R. L., & SHOEMAKER, D. J. Self acceptability and descriptions of sociometric choices. *J. abnorm. soc. Psychol.,* 1955, **51,** 260–262.

MANIS, M. Social interaction and the self concept. *J. abnorm. soc. Psychol.,* 1955, **51,** 362–370.

MANNHEIM, BETTY F. The influence of reference groups and membership groups on the self-image. Urbana, Ill.: Group Effectiveness Res. Lab., Univer. of Illinois, *Interim Tech. Rep. No. 5,* 1957.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana, Ill.: Univer. of Illinois Press, 1957.

PRECKER, J. A. Similarity of valuings as a factor in selection of peers and near-authority figures. *J. abnorm. soc. Psychol.,* 1952, **47,** 406–414.

ROGERS, C. R. Perceptual reorganization in client-centered therapy. In R. R. Blake & G. V. Ramsey. *Perception: An approach to personality.* New York: Ronald Press, 1951. Pp. 307–327.

ROGERS, C. R. Studies in client-centered psychotherapy. III: The case of Mrs. Oak—a research analysis. *Psychol. Serv. Cent. J.,* 1951, **3** (1–2), 47–165.

ROGERS, C. R., & DYMOND, ROSALIND F. *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.

ROSENTHAL, IRENE. A factor analysis of anxiety variables. Unpublished doctoral thesis, Univer. of Illinois, 1955.

STEINER, I. D., & PETERS, S. Conformity and the ABX system. *J. Pers.,* 1958, **26,** 229–242.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48,** 285–290.

WILLOUGHBY, R. R. Neuroticism in marriage. The incidence of neuroticism. *J. soc. Psychol.,* 1934, **5,** 467–499.

WITMER, H., & KOTINSKY, R. (Eds.) *Personality in the making.* New York: Harper, 1952.

(Accepted for publication March 17, 1958)

APPENDIX A

INSTRUCTIONS AND QUESTIONNAIRES FOR THE DORMITORY GROUPS

This study is designed to investigate the way in which people think about themselves and about those with whom they live. The results which are obtained will be completely confidential and will not affect you in any way. With your cooperation, we hope to discover some of the factors which are important in group life. You can contribute to this study by giving your frank and immediate reactions to the items which are presented on the following pages.

Each of these items will consist of two words which are opposite in meaning, such as:

Confident ______ : ______ : ______ : ______ : ______ : ______ : ______ Unsure

Note that these words are separated by seven spaces. Your task will be to use items of this type to describe yourself and several of your neighbors in the dormitory.

For example, if you were asked to describe yourself, you would put a check mark in the space which best tells how *confident* or *unsure* you ordinarily consider yourself to be. If you think of yourself as being *extremely confident* you would put a check mark in the space nearest the word *Confident,* as shown here:

Confident ___√___ : ______ : ______ : ______ : ______ : ______ : ______ Unsure

If you think of yourself as being *definitely confident,* but not *extremely confident,* you would place the mark one space farther away from the word *Confident,* like this:

Confident ______ : ___√___ : ______ : ______ : ______ : ______ : ______ Unsure

If you think you are *somewhat more confident than unsure* you would put your check mark on the *Confident* side of the middle.

Confident ______ : ______ : ___√___ : ______ : ______ : ______ : ______ Unsure

If you think you are *neither confident nor unsure* you would put your check mark in the middle space, like this.

Confident ______ : ______ : ______ : ___√___ : ______ : ______ : ______ Unsure

Similarly, if you think of yourself as being *somewhat more unsure than confident,* you would use the first space on the *Unsure* side of the middle.

Confident ______ : ______ : ______ : ______ : ___√___ : ______ : ______ Unsure

Finally, if you think of yourself as being *definitely unsure,* you would place your check one space away from the *Unsure* side of the scale, and if you think you are *extremely unsure,* you would use the space which is nearest to the word *Unsure.*

Please remember that there are no right or wrong answers. We want your first frank reaction to each of the items. You can be most helpful by working rapidly, without stopping to think much about your answers. Please do not omit any items and mark each item only once. Finally, remember to make each description in terms of the way you yourself think about the person, rather than in terms of his reputation.

Thank you for your cooperation.

SCALE SHEET

CONFIDENTIAL

Cooperative ______ : ______ : ______ : ______ : ______ : ______ : ______ Uncooperative

Quitting ______ : ______ : ______ : ______ : ______ : ______ : ______ Persevering

Stable ______ : ______ : ______ : ______ : ______ : ______ : ______ Unstable

Confident ______ : ______ : ______ : ______ : ______ : ______ : ______ Unsure

Seclusive ______ : ______ : ______ : ______ : ______ : ______ : ______ Sociable

Immature ______ : ______ : ______ : ______ : ______ : ______ : ______ Mature

Adventurous ______ : ______ : ______ : ______ : ______ : ______ : ______ Timid

Thankless ______ : ______ : ______ : ______ : ______ : ______ : ______ Grateful

Friendly ______ : ______ : ______ : ______ : ______ : ______ : ______ Hostile

Energetic ______ : ______ : ______ : ______ : ______ : ______ : ______ Subdued

Impatient ______ : ______ : ______ : ______ : ______ : ______ : ______ Patient

Softhearted ______ : ______ : ______ : ______ : ______ : ______ : ______ Hardhearted

Unreflectful ______ : ______ : ______ : ______ : ______ : ______ : ______ Thoughtful

Frank ______ : ______ : ______ : ______ : ______ : ______ : ______ Secretive

Meek ______ : ______ : ______ : ______ : ______ : ______ : ______ Forceful

Impulsive ______ : ______ : ______ : ______ : ______ : ______ : ______ Deliberate

Easygoing ______ : ______ : ______ : ______ : ______ : ______ : ______ Short-tempered

Practical ______ : ______ : ______ : ______ : ______ : ______ : ______ Impractical

Boastful ______ : ______ : ______ : ______ : ______ : ______ : ______ Modest

Intelligent ______ : ______ : ______ : ______ : ______ : ______ : ______ Unintelligent

Depressed ______ : ______ : ______ : ______ : ______ : ______ : ______ Cheerful

Reliable ______ : ______ : ______ : ______ : ______ : ______ : ______ Undependable

Unrealistic ______ : ______ : ______ : ______ : ______ : ______ : ______ Realistic

Much interest in opposite sex ______ : ______ : ______ : ______ : ______ : ______ : ______ Little interest in opposite sex

CONFIDENTIAL

APPENDIX B

INSTRUCTIONS AND QUESTIONNAIRES FOR THE BARRACKS GROUPS*

People differ in the ways they think about themselves and about those who are most important to them. Please give your immediate, first reaction to the items on the following pages.

On each sheet are pairs of words which are opposite in meaning, such as *Talkative* and *Quiet*. You are asked to describe yourself and each of several people who are important to you by placing a check in one of the six spaces on the line between the two words.

Each space represents how well the adjective fits the person you are describing, as if it were written:

Talkative __________ : __________ : __________ : __________ : __________ : __________ Quiet

very talkative	quite talkative	more talkative than quiet	more quiet than talkative	quite quiet	very quiet

FOR EXAMPLE: If you were to describe yourself, and you ordinarily think of yourself as being *quite talkative,* you would put a check in the second space from the word *Talkative,* like this:

Talkative __________ : ____✓____ : __________ : __________ : __________ : __________ Quiet

If you ordinarily think of yourself as somewhat *more quiet than talkative,* you would put your check on the quiet side of the middle.

Talkative __________ : __________ : __________ : ____✓____ : __________ : __________ Quiet

If you would think of yourself as *very quiet,* you would use the space nearest the word *Quiet.*

Talkative __________ : __________ : __________ : __________ : __________ : ____✓____ Quiet

Look at the words at both ends of the line before you put in your check mark. Please remember that there are no *right or wrong answers.* Work rapidly; your first answer is likely to be the best. Please do not omit any items and mark each item only once.

---

* **Tank and anti-aircraft artillery crews were given similar instruction and identical scale sheets.**

SCALE SHEET

First describe yourself as you ordinarily think about yourself.

MYSELF AS I AM

Friendly ______ : ______ : ______ : ______ : ______ : ______ Unfriendly

Cooperative ______ : ______ : ______ : ______ : ______ : ______ Uncooperative

Quits easily ______ : ______ : ______ : ______ : ______ : ______ Keeps trying

Calm ______ : ______ : ______ : ______ : ______ : ______ Upset

Confident ______ : ______ : ______ : ______ : ______ : ______ Unsure

Immature ______ : ______ : ______ : ______ : ______ : ______ Mature

Bold ______ : ______ : ______ : ______ : ______ : ______ Timid

Ungrateful ______ : ______ : ______ : ______ : ______ : ______ Grateful

Energetic ______ : ______ : ______ : ______ : ______ : ______ Tired

Impatient ______ : ______ : ______ : ______ : ______ : ______ Patient

Thoughtless ______ : ______ : ______ : ______ : ______ : ______ Thoughtful

Frank ______ : ______ : ______ : ______ : ______ : ______ Secretive

Careless ______ : ______ : ______ : ______ : ______ : ______ Careful

Easygoing ______ : ______ : ______ : ______ : ______ : ______ Quick-tempered

Practical ______ : ______ : ______ : ______ : ______ : ______ Impractical

Boastful ______ : ______ : ______ : ______ : ______ : ______ Modest

Intelligent ______ : ______ : ______ : ______ : ______ : ______ Unintelligent

Gloomy ______ : ______ : ______ : ______ : ______ : ______ Cheerful

Responsible ______ : ______ : ______ : ______ : ______ : ______ Undependable

Efficient ______ : ______ : ______ : ______ : ______ : ______ Inefficient

SCALE SHEET

Now describe yourself as you would most like to be.

MYSELF AS I WOULD MOST LIKE TO BE

How we think of others is important also. On this sheet describe your roommate.

HIS NAME____

On this sheet describe one of the other two people in your suite.

HIS NAME____

On this sheet describe the last member of your suite.

HIS NAME____

If you had a personal problem which you wanted to talk over with someone, to whom in your own or a neighboring building would you most likely go?

HIS NAME____

Now describe this person, unless you already have.

Does anyone in the building come to you with his problems? If so, will you name and describe this person?

HIS NAME____

How we think of our co-workers is also important. Think of someone with whom you have been able to work *best;* a man with whom you might get a job done easily and well, and describe him on this form. This may be someone you have known in the past, or someone you know now.

Everybody can work better with some people than with others. Think of someone with whom you have been able to work *least well,* a man with whom you might have difficulty in doing a job. Describe him.

What other people think of us is also important. Try to describe yourself as others in your building probably see you.

MYSELF AS OTHERS IN MY BUILDING PROBABLY SEE ME

Vol. 73, No. 4 Whole No. 474, 1959

Psychological Monographs: General and Applied

THE INFLUENCE OF DIFFERENT ACTIVITIES ON THE APPARENT LENGTH OF TIME[1]

JOHN C. LOEHLIN[2]

University of California

IT IS A COMMONPLACE of daily experience that time seems to pass more rapidly under some conditions than it does under others. And yet such familiar phenomena as the slow passage of time in boredom, depression, or suspense, or its rapid flow when the attention is pleasantly absorbed, have not been greatly clarified at the hands of psychologists.

One reason for this may well be that the experimental manipulation of variables of attention and attitude is uncertain at best. This situation is aggravated by the fact that many investigators of temporal perception have preferred to work with intervals only a few seconds in length, which offer many advantages in terms of psychophysical elegance but little scope for the manipulation of mood.

On the other hand, the investigators who have worked with longer intervals, although they have amply confirmed that time will appear to differ in its rate of passage under different conditions, have varied so widely in the conditions they have chosen that it is by no means certain whether the concepts invoked by one are applicable to the results of another. A few examples may serve to make this clear.

Harton (1938) found that subjects making difficult weight discriminations estimated the time to be shorter than during an equal period spent making easy discriminations. Axel (1924) found that subjects estimated a period spent solving analogies as shorter than an equal period spent canceling 5's. Rosenzweig and Koht (1933) found that subjects estimated a "test" period with an insoluble puzzle as shorter than an equal "practice" period. On the face of it, one might suspect that these three results had a considerable amount in common, but the three authors offer quite different concepts to interpret them. Axel speaks of higher and lower levels of behavior, Harton of unity of organization, and Rosenzweig and Koht of need-tension.

Other experimenters have had their subjects estimate the time spent taking dictation, listening to a person read, listening to music and metronomes, canceling letters, stringing beads, submitting to pain, noise, and electric shock, solving puzzles and mazes, making public speeches, reading aloud, filling out questionnaires, and just resting. And in addition to the concepts mentioned above, they have, in interpreting their results, used such terms as filled and unfilled time, concentration, level of attention, kinesthetic strain, satiation, stress, self-organization, attractiveness of or distance to a goal, and so forth.

It seems unlikely that all these concepts are necessary to interpret the experimental findings. And yet it seems doubtful that any one of them is sufficient. The first task of the present investigation was, therefore, to determine how many and what kinds of variables are necessary in order to account for variations in the apparent length of time, in situations of the kind which experimenters in this area have traditionally been concerned with.

---

[1] This paper is a portion of a dissertation submitted to the Department of Psychology of the University of California in partial fulfillment of requirements for the degree of Doctor of Philosophy.

It is with pleasure that I acknowledge the helpful counsel of my thesis director, D. W. MacKinnon, and the other members of my dissertation committee, C. F. Wrigley and W. R. Dennes. In addition, I owe a particular debt of gratitude to R. S. Crutchfield for valuable discussions during the formulation of the project.

For making possible the electronic computations, I am grateful to the Operating Committee and staff of the University of California Computer Center, and to the National Science Foundation, which contributes to the support of the Center. I would also like to express my thanks to Jay Adams, J. O. Neuhaus, and William Murr for their advice and assistance in connection with the computations.

[2] Now at the University of Nebraska.

Given such variables, further questions arise concerning their generality. Are they equally important for various lengths of interval? It seems likely that some influences on time perception in longer periods are equally effective during short intervals, and that others are not. Different methods have been used in experiments in time perception: To what extent are the same variables involved regardless of the method of measurement? And finally the question always arises: How important are such experimentally derived variables in dealing with time perception outside the laboratory? The present investigation did not attempt to deal comprehensively with all these questions, but it did endeavor to gain at least preliminary information along each of these lines.

Design of the Experiment

The first goal of the present research was to determine what different kinds of influence the contents of time intervals have on the apparent duration of the intervals. One sort of evidence for believing that different situations are similar or different in their effect on time estimation is the degree of correlation between persons' time estimates in the situations. If individuals are similarly ordered in Situations A and B in terms of the length of their time estimates, but differently ordered in Situation C, one has some grounds for supposing that A and B are affecting the estimates in the same way, and that C is having a different influence. Given a number of such situations, and the corresponding correlations, it should be possible to gain some notion as to how many different influences are operating, and what kinds. In the present investigation, the technique of factor analysis seemed the obvious method to use for this purpose.

Since, in such an analysis, the factors obtained are ultimately derived from the way situations affect persons, they may reflect both characteristics of the situations and characteristics of the persons. In some uses of the factor analytic technique, interest is focused primarily on one or the other of these. The literature of time perception suggests, however, that in this area we may be dealing quite directly with differences in the way situations affect persons. Therefore, in the present design, provision is made for varying the situations in which time is judged, for evaluating some characteristics of the persons doing the judging, and for determining several aspects of the persons' perceptions of the situations.

It should be noted that an approach of this kind depends on individual differences. A variable that affected all persons in exactly the same way would not influence the correlations between situations, and hence could not emerge as a factor in the analysis. While a complete absence of individual differences is hardly to be expected, tendencies toward greater or less uniformity presumably exist, and could well affect the weight certain variables have in a factor analysis—a possibility to be kept in mind in interpreting the results.

The 2-Minute Tasks

The basic part of the experiment was the obtaining of estimates of the duration of a 2-minute time interval occupied in a variety of ways. The conditions were selected with several criteria in mind, including the following: (*a*) they should provide a good sampling of the kinds of conditions known from earlier research to affect time estimates; (*b*) they should afford a minimum of irrelevant cues to length—the *S*s should be given as few grounds as possible for comparing different tasks on such bases as the proportion of the task accomplished, number of items done, etc.; (*c*) some periods should be included which the *S*s would not expect to estimate; (*d*) some of the tasks should provide scores on other dimensions which might be relevant to time estimation; and (*e*) the tasks should be suitable for group administration.

The following conditions were selected. They are listed in the order in which they were administered.[3]

1. *Introduction.* A 2-minute period occupied by *E*'s introduction of the experiment, the distribution of test booklets, filling in of face-sheet data, etc.
2. *Writing 1's.* The *S* wrote the figure "1" repeatedly on a blank sheet of paper.
3. *Easy anagram.* The *S* made words out of letters from a given long word (TRANSPORTATION).
4. *Difficult comparisons.* The *S* made psychophysical judgments—comparisons of pairs of short time intervals.
5. *Gottschaldt practice.* The *S* solved hidden-figure puzzles "for practice."
6. *Gottschaldt test.* The *S* solved similar puzzles, as a "test."
7. *Slow calculation.* Cumulative mental addition and subtraction of numbers from 1 to 5, called out at a slow rate by *E*.
8. *Listening to reading.* The *S* listened while *E* read aloud a passage of humorous prose.
9. *Pleasant thoughts.* The *S* was instructed to close his eyes, relax, and think about any pleasant topic.
10. *Interruption.* An apparent apparatus breakdown, which required two minutes for *E* to "repair."
11. *Writing 1's.* Repetition of No. 2 above.
12. *Easy comparisons.* Repetition of No. 4, but quite easy comparisons.
13. *Difficult anagram.* Repetition of No. 3, but a more difficult word (MISSISSIPPI).
14. *Counting "the's."* The *S* counted occurrences of the word "the" in another prose passage read aloud by *E*.
15. *Concentration on time.* The *S* was instructed to close his eyes and concentrate on the experience of the passage of time.
16. *Questionnaire.* The *S* was occupied in responding to the items of a questionnaire concerning his everyday time perception.

The activities were all of the same length—two minutes—to permit direct comparisons among them. In preliminary tests it appeared that the variation of the apparent length of the intervals was sufficiently great that almost none of the *S*s had any suspicion of their objective equality.

The activities were administered in the same order to all *S*s: first, for considerations of experimental convenience, and second, because it was undesirable to change the positions of some of the items—e.g., *Concentration on time* had to be given toward the end to avoid possible contaminating effects on the other estimates, and *Introduction* was of necessity given at the start. The fixed order did, however, introduce certain complications (see the interpretation of Factors I and III below).

The Nature of the Estimates

In previous studies of intervals of moderate length, a direct estimate in conventional time units has usually been used. As contrasted with comparison or reproduction methods, this method has the advantage that the estimates are both more readily made and more simply interpreted. With techniques in which two intervals are compared it is always a problem to determine which of the intervals is being affected in a particular instance—if the second is judged the longer of the two, is it because it is being overestimated, or because the first is being underestimated, or because both things are happening, or because both intervals are being over- or underestimated to different degrees? Decision among these alternatives often presents thorny problems. The main disadvantages of the direct estimate method, on the other hand, are (*a*) the fact that the distribution of estimates tends to be skewed by the proximity of the zero point, particularly if the intervals are short, and (*b*) the tendency of many subjects to give "round number" estimates—to favor estimates of whole minutes and simple fractions of a minute.

Since neither of these tendencies, in a moderate degree at least, would be too serious in the present investigation, the method of estimation in conventional time units was adopted. The *S*s were asked to make their estimates to the nearest ten seconds. With the expected range of judgments, this would provide satisfactory discrimination, and it was hoped that this unit would be large enough for *S*s to actually use, whereas asking for an estimate to the nearest second, as is usually done, almost inevitably drives *S* to a higher-order unit such as the quarter- or half-minute.[4]

---

[3] A detailed description of the experimental procedures, including instructions to *S*s, may be found in Loehlin (1957, pp. 45–58).

In previous investigations using the method of direct estimation, *S*s have usually simply been asked to estimate how long a given interval was. But this presents problems. There are a variety of possible determinants of the estimate *S* makes of a time interval, and how long the interval seems to *S* is only one. Assuming that the experimental arrangement is successful in avoiding gross give-away cues, the most important additional variable will probably be *S*'s belief as to whether, under the given conditions, time is likely to seem longer or shorter to him than it really is. During preliminary experiments in the present investigation, it became clearly apparent that many *S*s will, in fact, make allowances of this kind when asked to estimate the length of intervals of time. If the interest of the experimenter is simply in how well *S* can estimate time under various conditions, this presents no problem. But if his concern is with how the conditions affect *S*'s perception of the time, the matter must be dealt with.

The solution adopted was to have *S* make *two* estimates—one of how long the time seemed, and one of how long he thought it actually was. The procedure is somewhat analogous to the "perceptual" and "critical" attitudes used by Brunswik (1949) and others in the study of constancies and other perceptual phenomena. It was hoped that *S*'s judgment of how long the time seemed would follow fairly directly from his perception of it, while any deliberate inferences based on his knowledge of the situation would be confined to his estimate of how long the time actually was. The estimate of how long the time seemed was always made first and put in a "box" on the response sheet. The estimate of how long the time actually was was made afterwards and written outside the box to emphasize its derived character. The *S*s were not forced to make a distinction; the instructions permitted both estimates to be the same. For a few *S*s the distinction did seem to be a meaningless one.[5]

After making the two estimates, *S*s rated each task on five 5-point scales. They indicated whether the activity seemed interesting or boring, difficult or easy, satisfying or frustrating; whether the time seemed to pass slowly or rapidly; and whether they wished the period had been shorter or longer.

Other Time Judgments

Along with the judgment of 2-minute intervals, some estimates of shorter and longer intervals of time were included, and some comparisons of short time intervals. The main purpose of these additional measures was to gain generality—to determine whether the ways in which content influenced the estimation of the 2-minute intervals would affect a wider range of times, and whether these influences might also contribute to certain other phenomena in the area of time perception, namely, time errors and other illusions occurring in the comparison of short time intervals.

Two short periods of time were included: one second and four seconds. Four estimates were made of each of these two intervals, the estimates being spaced among the other procedures. The *S*s were instructed to make their estimates to the nearest tenth of a second. No distinction was made between how long time seemed and how long it was on the assumption that this distinction would not be very meaningful for such short periods, particularly the 1-second interval.

To provide an estimate of a longer interval, a 20-minute period was used (the first 20 minutes of the experimental session). This particular interval was selected as one that was long enough to be clearly of a different order from the 2-minute period, without being so long as to approach the larger known interval in which it was embedded, the 50-minute class hour. The *S*s made estimates both of how long the interval seemed, and how long they thought it actually was.

---

[4] This hope was only partially realized. A good many *S*s used higher-order units anyway.

[5] For some evidence suggesting that the procedure had the effect desired, see Loehlin (1957, pp. 72–78).

Two procedures in which short intervals were compared were also chosen for inclusion. In one of these, pairs of empty intervals were compared in order to measure the time error. Eighteen trials were used; in each a standard 4-second interval was presented first, followed by a comparison interval which varied from $2\frac{1}{2}$ seconds to $4\frac{1}{2}$ seconds, with the largest number of trials in the neighborhood of $3\frac{1}{2}$ seconds. (A negative time error was anticipated.)

For the other comparison procedure, one was chosen in which the two compared intervals actually differed in their content. The technique used by the writer in an earlier, unpublished investigation was employed. An accelerating pattern of four sounds was compared with a decelerating pattern (a comparison of the form .   . . .). Each pattern was of the same total length — again four seconds — and twelve comparisons were made, half in each order.

Time Questionnaire

To provide the possibility of generalization of a different kind, *S*s were asked at the end of the experiment to fill out a questionnaire consisting of 85 true–false, self-descriptive items concerning their perception of and attitudes toward time in everyday life situations. (The questionnaire is reproduced in Loehlin [1957] as Appendix F.) The items for this questionnaire were mostly newly written, although a few were adapted from other sources, such as a questionnaire used by Adler (1954). An effort was made to include a variety of items from everyday experience, as well as items derived from aberrations of behavior and perception concerning time as reported in the clinical literature (e.g., Bonaparte, 1940; Dooley, 1941; DuBois, 1954; Schilder, 1936).

Other Measures

Several of the experimental procedures permitted the derivation of additional scores. It seemed desirable to represent at least two general areas of personal function. One was the individual's characteristic flow of activity, as represented by variables such as fluency, personal tempo, and the like. The other was his ability to perceive accurately in spite of distraction: to analyze out aspects of a complex perceptual situation.

For the area of fluency, scores were obtained for (*a*) the number of 1's written in two 2-minute periods (under instructions to write at a natural rate); (*b*) the number of solutions to the easy anagram; and (*c*) the number of questionnaire items completed in a 2-minute interval.

For measurements of the ability to perceive analytically, scores were obtained for (*a*) the Gottschaldt figures and (*b*) the "counting the's" procedure. In both these tasks, *S* must be able to isolate highly embedded features of the situation and to resist the distraction of irrelevant aspects.

The MMPI

Finally, as an instrument to evaluate differences of an emotional and motivational nature between individuals, the Minnesota Multiphasic Personality Inventory (MMPI) was chosen. The MMPI's were completed on a voluntary basis at a date several months after the main experiment, and were available for about 70% of the subjects.

Summary

To summarize briefly, the design of the experiment was as follows: (*a*) Obtain from a group of *S*s their estimates of how long time seems in a variety of situations, as well as estimates of different lengths of interval, measures of other time perception phenomena, and scores on other related variables. (*b*) Intercorrelate these measures and factor-analyze. (*c*) Interpret the factors thus obtained using knowledge of the situations, knowledge of the *S*s' ratings of the situations along various dimensions, and knowledge of the personality organization of the subjects. And finally, (*d*) consider the relationship of these factors to the findings of previous investigators in this area, and to the *S*s' reports of their

TABLE 1

NUMBERS OF *S*S, BY AGE, SEX, AND PSYCHOLOGY COURSE

Course	Age						All Ages
	17	18	19	20	21	22	
Introductory							
Men	—	1	5	3	1	—	10
Women	—	7	13	4	1	—	25
Personality							
Men	—	1	3	4	2	1	11
Women	1	8	24	9	3	1	46
Industrial							
Men	—	—	—	1	8	1	10
Women	—	—	—	—	3	—	3
Totals	1	17	45	21	18	3	105

behavior, perception, and attitudes with respect to time.

COLLECTION AND ANALYSIS OF THE DATA[6]

Subjects

The *S*s were 105 members of undergraduate psychology courses at the University of California at Berkeley, who participated in the experiment as part of a course requirement. They were obtained from the introductory psychology course, from a lower-division course in personal and social adjustment designed for non-majors in psychology, and from an upper-division course in industrial psychology. Psychology majors in the last-named class were excluded from the sample.

In view of the possibility that large variations in age may affect time estimation (Adler, 1954), *S*s falling outside the range of 17-22 years were eliminated from the sample. (Swift and McGeoch [1925] found no relationship of age to time estimation in their experiment, with an age range approximately the same as that used in the present study.) The distribution of the subjects according to age, sex, and psychology course is shown in Table 1.

The *S*s were tested in small groups, in a two-hour experimental session. The size of the groups varied from one to nine persons, depending on the vagaries of sign-ups, with about 80% of the total sample being tested in groups of four to eight persons.

Scoring

The following 28 measures were selected for inclusion in a matrix to be intercorrelated and factor-analyzed:

1–16. Estimates of the sixteen 2-minute tasks. ("How long time seemed.")

17. Estimate of 1 second—mean of Trials 2–4. ($r_{1\text{I}} = .79$)[7]

18. Estimate of 4 seconds—mean of Trials 2–4. ($r_{1\text{I}} = .84$)

19. Estimate of 20 minutes.

20. Negative time error, comparison of 4-second intervals—number of trials on which the second interval was judged longer, in *Difficult comparisons.* ($r_{1\text{I}} = .50$)

21. Amount of illusion, accelerating and decelerating patterns—number of trials on which the decelerating pattern was judged longer. ($r_{1\text{I}} = .44$)

22. Number of 1's written—two trials combined. ($r_{1\text{I}} = .82$)

23. Number of questionnaire items completed in two minutes.

24. Number of *Easy anagram* solutions in two minutes.

25. Number of *Gottschaldt figures* solutions in four minutes—practice and test periods combined. ($r_{1\text{I}} = .83$)

26. Error, *Counting "the's"*—discrepancy between *S*'s answer and the correct score.

27. Number of persons in *S*'s test group.

28. *S*'s age in years.

Factor Analysis

The scores were punched on IBM cards, and with the program of Beckman and Quarles (1956) means and standard deviations for all the variables (Appendix A), the matrix of Pearsonian *r*'s of each variable with each other (Appendix B), and the inverse of that matrix, were computed by the IBM 701 electronic computer.

From the inverse matrix, the square of the multiple correlation of each variable with all the remaining variables was calculated, providing an estimate of the communality. Guttman (1954b) has shown that the exact communality will never be lower than this value, and Wrigley (1956) has found the squared multiple correlation to be a good initial estimate of the communality, in an empirical comparison with a number of other methods of estimation.

---

[6] A more complete description of the treatment of the data may be found in Loehlin (1957, pp. 59–71).

[7] All reliabilities given are split-half, corrected by the Spearman-Brown formula.

The correlations and estimated communalities (to three decimal places) were then repunched on IBM cards and a principal axes factor analysis was computed by the IBM 701, using a program written by J. O. Neuhaus. The loadings for the twelve factors of greatest variance are given as Appendix C.

In choosing the number of factors to rotate, indications from several of the empirical and theoretical criteria cited in the literature were considered (see, e.g., Cattell, 1952, p. 298 ff). Most of these criteria were originally developed for the purpose of determining when to stop extracting factors when using the centroid methods, but similar considerations apply in the present situation. As is typical, the different criteria gave somewhat different indications. McNemar's criterion, which "tends in general experience to stop factorization too early" (Cattell, 1952, p. 300), dropped below its critical value after five factors. The Guilford-Lacey rule (Thomson, 1951, p. 122) gave its first value below the critical one for the fifth factor, but the sixth and ninth factors again exceeded the criterion. Mosier's requirement (Cattell, 1952, p. 299) of a leveling off of the standard deviations of successive sets of residuals might be taken to indicate the insignificance of all factors after the first, the third, the eighth, or some subsequent factor, depending on how narrowly it is applied. Saunders' criterion, which Cattell describes as probably the "theoretically and practically most effective" of the eleven he lists (Cattell, 1952, p. 301), did not pass its critical value until after nine factors.

After some exploratory rotations with a few factors, the first eight factors were selected for rotation. A graphical technique was used until the final position of the axes was approached, when a rotating matrix was computed and accurate loadings obtained algebraically. Then final graphical adjustments were made, the final rotating matrix derived, and the final matrix of factor loadings computed. (See Appendices C–E for the original factor matrix, the rotating matrix, and the final factor matrix.) The ninth and tenth factors were plotted against Factors I–VIII following rotation, to see if their inclusion in further rotation were merited. Factor IX failed to suggest any additional rotations; Factor X suggested only one (with Factor IV), and the indications for that one were somewhat doubtful. The addition of two dimensions of questionable significance in order to secure a possible improvement in the position of one axis did not seem justified.

The axes were kept orthogonal during rotation, not so much on the basis of theoretical preconceptions, as in the absence of any very strong indications for the abandonment of orthogonality. The first factor could reasonably be placed oblique to some of the rest; on the other hand, its retention in its present form, as a factor general to the 2-minute estimates, seems psychologically meaningful. The remaining factors show no indications of more than very slight obliquity among themselves.

The structure resulting from the present rotation meets all but one of Thurstone's (1935) original criteria of simple structure (considering a loading of less than .10 to be negligible [Thurstone, 1935, p. 156]). The deficiency is in an insufficient number of zero-order loadings on the first factor. The small number of variables in the hyperplane of this factor is attributable to the large number of 2-minute estimates in the matrix, since Factor I is general to these. The present structure is also satisfactory, again excepting Factor I, by Cattell's criterion of hyperplane area (1/3 to 1/2 the variables on each factor in the plus-or-minus .10 hyperplane [Cattell, 1952, p. 238]).

Factor Scores

A tentative interpretation of the rotated factors suggested that four (Factors I, II, IV, and V) would be of principal interest from the standpoint of time perception. Approximate factor scores were derived for these as follows:

Factor I. Two-minute estimates: *Difficult comparisons* + *Gottschaldt practice* + *Pleasant thoughts* + *Questionnaire.*

Factor II. 4 × estimate of 1 second + estimate of 4 seconds.

Factor IV. Two-minute estimates: *(Writing 1's (2nd) + Hard anagram) − (Writing 1's (1st) + Easy anagram)* + 50.

Factor V. (Negative time error, *Difficult comparisons* + Amount of illusion, *Accel.-decel. patterns*) reflected.

Intercorrelations of the four factor scores were computed, and the scores proved to be fairly independent, as desired. The only r's greater than ± .05 were $r_{\mathrm{I\,II}} = .14$ and $r_{\mathrm{IV\,V}} = -.16$.

Persons scoring in the highest and lowest 27% on each of these factors were selected for comparison of their ratings of the various activities and their MMPI scores. Each of these high- and low-scoring groups contained 28 persons.

MMPI

Five of the standard MMPI clinical scales, *D, Hy, Pd, Pt,* and *Ma,* were selected as being the most likely to be related to differences in time perception, and to these was added the Ego Strength scale developed by Barron (1953). The MMPI's of the 74 *S*s who took the test were scored on these scales in the standard manner (Hathaway & McKinley, 1946), including the *K*-correction. The validation scores were examined, and one *S* who obtained an *F* score of 17 was dropped from the sample, leaving 73 subjects.

For the groups of 28 high and low *S*s on each of the factors, MMPI's were available as indicated in Table 2. These differences are of an order attributable to chance. Thus, while those *S*s for whom MMPI's were not obtained may have differed in some ways from those who took the test, it does not appear that these differences are significantly related to their time perception scores.

TABLE 2

NUMBER OF MMPI'S AVAILABLE, FOR *S*S WITH HIGH AND LOW FACTOR SCORES

(28 *S*s Per Group)

Group	Factor			
	I	II	IV	V
High scorers	20	21	20	16
Low scorers	19	16	21	21

RESULTS[8]

Interpretation of the Factors

Factors I and III. (See Tables 3 and 4.) Factors I and III have most of their load-

TABLE 3

VARIABLES LOADED ON FACTOR I

No.	Variable	Loading
9	2-min. est., Pleasant thoughts	.70
13	2-min. est., Hard anagram	.67
14	2-min. est., Counting "the's"	.64
12	2-min. est., Easy comparisons	.59
15	2-min. est., Concentration on time	.58
7	2-min. est., Slow calculation	.56
10	2-min. est., Interruption	.55
16	2-min. est., Questionnaire	.52
5	2-min. est., Gottschaldt practice	.51
4	2-min. est., Difficult comparisons	.50
8	2-min. est., Listening to reading	.49
3	2-min. est., Easy anagram	.48
6	2-min. est., Gottschaldt test	.41
11	2-min. est., Writing 1's (2nd)	.40
2	2-min. est., Writing 1's (1st)	.26
23	No. of questionnaire items in 2 min.	−.35
24	Easy anagram solutions	−.40

TABLE 4

VARIABLES LOADED ON FACTOR III

No.	Variable	Loading
6	2-min. est., Gottschaldt test	.63
5	2-min. est., Gottschaldt practice	.58
19	20-min. estimate	.46
4	2-min. est., Difficult comparisons	.38
2	2-min. est., Writing 1's (1st)	.38
3	2-min. est., Easy anagram	.34
1	2-min. est., Introduction	.32
23	No. of questionnaire items in 2 min.	.27
25	Gottschaldt figures solutions	−.28

---

[8] For a presentation of the results of the separate procedures, see Loehlin (1957, pp. 72–83).

ings on the 2-minute estimates, and all the 2-minute estimates have substantial loadings on one or both of these factors. The factors reflect the general positive intercorrelation of the 2-minute estimates, and appear to have resulted from consistent individual differences in such estimates.

However, before inquiring into the possible nature of such individual differences, it is necessary to ask why there are *two* factors involved rather than only one. The answer appears to lie in the fact that estimates made close together in the course of the experiment tend to correlate more highly than estimates more widely separated. This can readily be seen in the intercorrelations of the 2-minute time estimates (Appendix B): the correlations tend to be highest next to the principal diagonal, and to drop off toward the upper right-hand and lower left-hand corners. This form of correlation matrix has been described by Guttman (1954a) under the name of "simplex."[9] He notes that in a conventional factor analysis such a matrix tends to yield two factors, based on the clustering together of the earlier and later tests (Guttman, 1954a, p. 347). This indeed appears to have happened in our analysis. Factor I reflects the general intercorrelation of all the estimates, with the highest loadings on the later estimates in the series. Factor III's high loadings are on the estimates in the early part of the experiment (including the estimate of the first 20 minutes). Factors I and III have, therefore, been considered as a single factor, one which represents individual differences in the general level of estimates of the 2-minute interval. For convenience, it will be referred to hereafter simply as Factor I. The factor scores for "Factor I" are based on this combined factor.

It may be noted that such a factor need not be, and probably will not be, causally simple—various quite different conditions may jointly influence the over-all level of *S*'s estimates. General response tendencies—to overestimate time, give extreme responses, use round numbers, or the like—would be a possibility; however, these do not seem to be involved in a major way, for they ought to affect the short intervals as well, and overestimation of the short intervals does not load substantially on Factor I. However, any idiosyncratic notions *S*s may have as to how long a minute is would contribute to the intercorrelation among the 2-minute estimates, and hence to the loadings on this factor. The loading of the 20-minute estimate on this factor is consistent with this view.

Differences in ability, or practice in judging time intervals of this length, may also play a role since persons with low scores on this factor are closer to objective accuracy than persons with high scores. The negative correlations of Factors I and III with some of the performance measures suggest that ability may be involved to some extent, but the evidence for an ability factor is clearer in the case of Factor IV.

Another possible source of general intercorrelation of estimates would be differences between the experimental situation as a whole and the conditions under which *S* normally estimates time (assuming that these differences varied between individuals). Clues to such differences might lie in *S*'s ratings of the tasks, or in the personality measures.

As for the ratings, one difference[10] emerged: *S*s with high scores on Factor I tended to describe tasks loaded on the factor as less interesting than did *S*s with low scores. In other words, persons who overestimated these tasks tended also to rate them as less interesting. The first factor appears to reflect to some extent, then, individual differences in the degree of interest or involvement in the experiment as a whole.

---

[9] In the circumstances with which Guttman is primarily concerned, the higher correlation between adjacent variables arises in a different way. It is interesting to note, however, that one of the empirical examples he cites (Guttman, 1954a, p. 288) appears to resemble the present case at least as closely as it does his theoretical model.

[10] This difference is significant at the .05 level by *t* test; however, this and subsequent significance figures must be regarded with some caution, since a certain advantage may be taken of chance when a few *t* tests are selected from a larger number on the basis of their significance.

A second kind of difference which might contribute to a general intercorrelation of estimates would be an emotional reaction to the experimental situation as a whole, such as anxiety, annoyance, or apathy. Insofar as such a reaction was a characteristic one for the individual, it might be expected to be also reflected in a difference in MMPI scores, on such scales as *Pt, Pd,* or *D,* between individuals with high and low scores on Factor I. No significant differences of this nature emerged. However, since the MMPI was filled out at a time different from that of the experiment proper, transient differences in mood remain as a possible source of consistent individual differences over the set of 2-minute estimates.

To summarize, Factor I is general to the 2-minute estimates. It appears to reflect the tendency of some individuals to estimate the 2-minute intervals as consistently longer or shorter than others do. To some extent this may be due to individual differences in the use of the unit of estimate (minute). A second determinant seems to be the degree of interest in the tasks, with the subjects who find the tasks generally boring tending to estimate them as longer.

Factor II. (See Table 5.) This factor appears to be determined primarily by the estimates of the short intervals, but its other loadings suggest possible further implications. The 2-minute tasks whose estimates load on Factor II were among the tasks most overestimated; the period of concentration on time was estimated longest of all.

TABLE 5

Variables Loaded on Factor II

No.	Variable	Loading
18	Estimate of 4 sec.	.78
17	Estimate of 1 sec.	.70
15	2-min. est., Concentration on time	.36
2	2-min. est., Writing 1's (1st)	.27
11	2-min. est., Writing 1's (2nd)	.26
22	No. of 1's written in 4 sec.	−.29

Since the short intervals were "empty" ones, and since the tasks of concentrating on the passage of time and writing the figure "1" repeatedly are conspicuously lacking in variety of content, Factor II could perhaps be described as an "overestimation of empty intervals" factor. However, it is possible that attention to the passage of time is the significant variable, and that the relative emptiness of the intervals is important mainly because it permits this to occur. The primary role of the short intervals in determining this factor seems in accord with this interpretation. For a 1-second interval, emptiness does not suggest much in the way of emotional reactions such as boredom or suspense, but what the subjects *were* presumably doing during the short intervals was attending very closely to the passage of time.

The negative loading of the number of 1's written suggests that persons who "filled up" the empty intervals more with their activity showed less tendency to overestimate them. The ratings of the three 2-minute tasks by persons with high and low scores on Factor II seem consistent with this. The high scorers—persons who overestimated the short intervals—rated these three tasks as significantly less satisfying than did the low scorers ($p < .001$).

There were no substantial differences in MMPI scores between persons with high and low scores on Factor II, which may offer some additional weight to an interpretation of this factor in terms of differences in attention to time, rather than in terms of emotional reactions to empty time.

To summarize, therefore, it appears that Factor II, in addition to having variance specific to the short estimates, reflects the overestimation of empty intervals—possibly due to a greater attention to the passage of time during them.

TABLE 6

VARIABLES LOADED ON FACTOR IV

No.	Variable	Loading
13	2-min. est., Difficult anagram	.42
11	2-min. est., Writing 1's (2nd)	.41
1	2-min. est., Introduction	.33
23	No. of questionnaire items in 2 min.	.31
25	Gottschaldt figures solutions	.28
12	2-min. est., Easy comparisons	.27
26	Amount of error, counting "the's"	−.46

Factor IV. (See Table 6.) This factor appears to load two groups of variables: time estimates on repeated tasks, and scores reflecting good performance on tests involving the ability to perceive analytically.

The *Difficult anagram, Writing 1's (2nd),* and *Easy comparisons* tasks represent tasks given earlier in the series and now repeated (with changes in difficulty level for the anagrams and comparisons). Persons who tended to estimate the second administration as relatively long tended also to perform well on the *Counting "the's"* task and the *Gottschaldt figures,* as well as completing a large number of questionnaire items in two minutes.

The *Introduction* estimate seems on the face of it rather out of place among estimates reflecting repeated trials. This loading may, however, be an artifact—it may simply reflect the fact that the brighter *S*s finished filling out the face sheet data early, and then had to wait for the slower ones, and judged the period long in consequence. (Barring the few *S*s who completed the *Gottschaldt figures,* this was the only activity where the element of finishing early could enter: the others were all of indefinite length.)

There were no significant differences in the ratings which persons high and low on Factor IV gave to the tasks best measuring that factor, nor in the changes of ratings from the first to the second administration of these tasks. This tends to exclude a fairly obvious hypothesis—namely, that the brighter *S*s tended to judge a repeated task as long because they found it relatively less interesting.

Nor were there significant MMPI differences between *S*s with high and low scores on Factor IV. The largest difference (which approaches the .05 level) is for the high *S*s to score higher on Barron's Ego Strength scale. This, if reliable, is consistent with their better performance on the ability measures—Barron (1953) reports a positive correlation of this scale with several measures of intelligence.

To summarize, Factor IV appears to reflect a tendency to estimate the second of two periods of a repeated activity as relatively long. Positive loadings of perceptual ability measures, and a failure of ratings of the tasks to discriminate, suggest that this factor is of an intellectual rather than an attitudinal kind, but its precise nature remains uncertain.

Factor V. (See Table 7.) The highest loadings on this factor are on the two tasks in which *S*s compared short time intervals, and the loadings are negative, indicating that the persons with high scores on this factor tended to be those who were less affected by the illusions which characterize these tasks.

There is some reason to suspect that an activity–passivity dimension may be involved here. The *S*s with high scores on this factor tended to write a large number of 1's, and to overestimate the three periods in which *S* played a predominantly passive

TABLE 7

VARIABLES LOADED ON FACTOR V

No.	Variable	Loading
8	2-min. est., Listening to reading	.36
22	No. of 1's written in 4 min.	.32
1	2-min. est., Introduction	.31
27	No. of persons in *S*'s group	.29
19	Estimate of 20 min.	.28
10	2-min. est., Interruption	.26
3	2-min. est., Easy anagram	.25
20	Negative time error, Diff. comp.	−.43
21	Amount of illusion, Accel.-decel.	−.43

role *(Listening to reading, Introduction,* and *Interruption*). It seems plausible that *S*s preferring more active modes of judgment would be less subject to perceptual illusions than more passive *S*s. It is not obvious, however, why the estimate of 20 minutes should be loaded on this factor, and the positive loading of the *Easy anagram* estimate is, if anything, rather contradictory to an active–passive interpretation. The loading of number of persons on such a factor had not been anticipated, but it does seem conceivable that in the larger groups there was a sharper contrast between *E*'s active role and the group's passive one.

The *S*s high and low on Factor V were compared in terms of their ratings of the two tasks loading highest on this factor *(Listening to reading, Introduction).* The high *S*s tended to judge the tasks as easier than the low *S*s did ($p<.05$) and showed some tendency to rate the time as seeming to go more rapidly ($p<.10$).

The passive tasks are not overestimated, then, because they *seem* to be going more slowly for the active subjects. Possibly it is the difference from their usual attitude that is important: the change in level of involvement is greater for them when shifting from active to passive tasks than it is for the more generally passive subjects.

On an active–passive dimension, one would expect to find MMPI differences between high and low *S*s. One would expect the passive *S*s to be higher on the *Pt, Hy,* and *D* scales, and the active *S*s higher on *Pd* and *Ma.* One might also expect to find the active *S*s higher on Ego strength. The differences on *Pt, Hy, D,* and *Ma* are in the proper directions, with the first two near the .05 and .10 levels, respectively. The *Pd* and *Es* differences, although not large, are in the wrong direction for the hypothesis—the lows have slightly higher mean scores on these scales.

In general, then, although there is scattered evidence that Factor V reflects an active–passive dimension, its interpretation in these terms must remain tentative.

TABLE 8

VARIABLES LOADED ON FACTOR VI

No.	Variable	Loading
24	Easy anagram solutions	.33
7	2-min. est., Slow calculation	.30
23	No. of questionnaire items in 2 min.	.29
22	No. of 1's written in 4 min.	.28
19	Estimate of 20 min.	−.30
28	Age	−.44

Factor VI. (See Table 8.) It was suspected that this factor reflected sex differences, since age and sex were known to be correlated in the sample (the males tended to be older, largely as a result of sex differences in enrollment in the classes from which the *S*s were drawn. See Table 1.) The loadings on Factor VI were therefore compared with the correlations[11] between sex and each of the variables in the study. The r_{pb} between age and (female) sex was −.41, compared to the loading of −.44 of age on Factor VI. The other *r*'s tended to be low, but for each of the nine variables where either the factor loading or the r_{pb} was as high as .20, both the loading and the *r* were in the same direction. It seems reasonable to conclude, therefore, that sex differences were mainly responsible for the emergence of this factor. The factor does not seem to be of particular interest from the standpoint of time perception except in a negative sense, for only two time estimates are loaded on it, and neither is high. The positive loading of the scores on the two verbal tasks is consistent with the common finding of a slight female superiority in verbal ability.

Factor VII. (See Table 9.) Three of the five loadings on this factor involve the anagram tasks, and the factor may largely represent something fairly specific to those procedures. It does not seem advisable to attempt to give it any more general interpretation. It is of passing interest to note that both scores and estimates are loaded

---

[11] Point biserial *r*'s.

TABLE 9

Variables Loaded on Factor VII

No.	Variable	Loading
24	Easy anagram solutions	.43
13	2-min. est., Hard anagram	.26
2	2-min. est., Writing 1's (1st)	.26
3	2-min. est., Easy anagram	.25
14	2-min. est., Counting "the's"	−.39

in the same direction, i.e., *S*s who got higher scores also estimated the period as longer. It does not appear that this is in general true—the opposite effect occurred with the Gottschaldt figures; cf. Factor III (Table 4).

Factor VIII. (See Table 10.) This factor has loadings on a number of the 2-minute estimates, but its interpretation seems obscure. The tasks have little obviously in common either in their structure or in terms of *S*s' ratings of them. A factor score was computed, but high- and low-scoring *S*s did not differ significantly in terms of either task ratings or MMPI scores. The two highest loadings on Factor VIII are on estimates during the comparisons tasks, and variance specific to these tasks may have been instrumental in the emergence of this as a separate factor, but beyond that it will be left uninterpreted.

General comments. This interpretation of the eight rotated factors has left four which seem to be of particular relevance to time perception. One of the remaining factors was interpreted as an artifact due to constant presentation order, one as due to sex differences, one as mainly specific to the anagram tasks, and the fourth was not interpreted, although it appeared to load some variance specific to estimates of the psychophysical tasks.

TABLE 10

Variables Loaded on Factor VIII

No.	Variable	Loading
12	2-min. est., Easy comparisons	.44
4	2-min. est., Difficult comparisons	.35
11	2-min. est., Writing 1's (2nd)	.34
14	2-min. est., Counting "the's"	.30
19	Estimate of 20 min.	.30
7	2-min. est., Slow calculation	.29
8	2-min. est., Listening to reading	.26
16	2-min. est., Questionnaire	.25

One question should be raised concerning the four time perception factors. Three of them—Factors I, II, and V—appear to load variance specific to the long estimates, short estimates, and comparison tasks, respectively. However (except possibly for the first of these) each appears to reflect important common variance as well, and it is this which has mainly been taken into account in arriving at the interpretations given.

The important presence of specific variance is in some respects awkward theoretically. It need not challenge the meaningfulness—or the usefulness—of the present factors, but it does suggest the possibility of alternative solutions. One cannot be confident that exactly the same set of factors would have emerged from a factor analysis of the 2-minute estimates if an entirely different group of additional variables had been used.

The Questionnaire

The questionnaire items which correlated with each of the four time-perception factors at a significance level of .05 or better are listed below.[12] The items for each factor are listed approximately in decreasing order of significance.

Factor I. Persons scoring higher on this factor—those who in general overestimated the 2-minute intervals—agreed more often with the following items:

I have a one-track mind

I seem to be pressed for time more than most people

I have a poor memory for the past events in my life

---

[12] For an item with true and false responses equally divided, the biserial r which is just significant at the .05 level is .24. With more extreme divisions of responses it is higher.

I often find it hard to 'get going' on something

Often, although the days go slowly, the weeks and months seem to fly by

I haven't enough time to accomplish the things I most want to do

The future is too uncertain for a person to plan very far ahead

Sometimes I feel like I'm in a trance

I have a hard time remembering the dates things happened.

There were no items with which persons low on this factor agreed significantly more often.

The items which correlate with Factor I seem to relate to deficiencies in time perspective—to a failure to integrate the present with the past and future in the effective planning and carrying out of activities. The relationship is on the face of it a simple one: If the past and the future lack reality, the present activity looms larger.

A somewhat similar relationship has been reported in a factor-analytic study by Cattell and Gruen (1955), in which a number of time estimates were included in a battery of tests given to 11-year-old children. One of the resulting factors, which had its highest loadings on several time estimates, also loaded a measure of "unattainability of distant goals." (Other loadings on the factor were: poor arithmetic, errors in a complex reaction time experiment, and female sex.)

Factor II. Persons scoring higher on this factor—those who overestimated short intervals and monotonous tasks—agreed more often with the following items:

A period of a few minutes often seems like hours to me

I am often in a hurry

I have a hard time remembering the dates things happened

I daydream a good deal

I have many memories of childhood events that seem like they might have taken place only yesterday

When I wake up in the morning, it usually seems like I just went to sleep a few minutes before.

Persons scoring lower, agreed more often with:

My childhood days seem very unreal to me

My father is a rather dominant person

I sometimes think we would have been better off if clocks had never been invented

I usually have a pretty accurate idea what time it is.

The contrast here appears to be one between persons with an underdeveloped and an overdeveloped sense of the conventional scale of time. The high-scoring *S*s seem to view time in a much more flexible way: Minutes often "seem like hours"; childhood events seem like "only yesterday"; when they awake it seems like they have been asleep "only a few minutes." The low-scoring *S*s, on the other hand, say they usually have an accurate idea of time, but the psychological distance of childhood, the dominant father, and the resentment of clocks suggest oversocialization—that this was achieved at some cost.

It would seem indeed to follow that for the high-scoring *S*s there would be a great difference in the apparent length of time depending on whether or not its passage were being attended to, and for the low-scoring *S*s this effect would be minimal. However, the results are also consistent with other interpretations—one might expect many influences to have a greater effect on the high-scoring *S*s' perception of time.

Factor IV. Only one item was significantly related to this factor. Persons with higher scores on the factor—those who overestimated repeated tasks—tended to agree more often with the statement: I am easily bored. The implication of this is far from clear. They did not actually rate the tasks loading on this factor as more boring than the low-scoring *S*s did, nor did they differ in the change in their boredom ratings from an original to a repeated activity.

Perhaps the most relevant fact after all is that only the one item emerged from this analysis. This would appear to add weight to the view that Factor IV is largely independent of attitudinal and motivational factors, which are relatively well represented in the questionnaire, and is fairly specifically linked with the cognitive variables reflected in the performance scores.

Factor V. Persons scoring higher on this factor—those who were less subject to time illusions—agreed more often with the following items:

> I have a pretty definite idea what I will be doing next summer
> I usually have a pretty accurate idea what time it is
> I'm pretty good at saving time
> I have great faith in the future.

Persons scoring lower, agreed more often with these items:

> I become impatient if I have to wait five minutes for someone
> I have a hard time getting things done without deadlines.

The high items appear to reflect a feeling of confident control of time. The low end of the dimension is less positively defined, but both items suggest an individual at the mercy of external events.

This appears reasonably consistent with the active–passive interpretation suggested earlier for this factor.

Discussion and Conclusions

The major purpose of the investigation was to examine the nature of the influences that various activities have on the apparent length of time during them. A factor-analytic design was employed with the aim of identifying the main ways in which different activities influence the perception of moderate intervals of time.

The 2-Minute Intervals

What, then, proved to be the major variables related to the perception of moderate intervals of time in this investigation?

The first such variable seems to have been the interest or boredom which the tasks aroused in *S*s. This variable appears as a partial determinant of general under- or overestimation of time, as reflected in Factor I. Persons who found the tasks in general to be less interesting tended also to perceive them as longer.

The second variable is related to the overestimation of intervals lacking in variety of content. It appears that this might be due to greater attention given to the passage of time during such periods. The persons who overestimated these periods most showed evidence on the questionnaire that their perception of time in everyday situations is relatively fluid. The persons who overestimated these periods least indicated a better regulated sense of time in other situations. This regulation may in part be mediated by compensatory activity in boring situations—the latter *S*s wrote more 1's, and found the monotonous tasks relatively satisfying.

A third variable concerns the over- or underestimation of repeated activities. This was related to measures of perceptual ability: The more able *S*s tended to estimate repeated tasks as relatively longer, the less able *S*s as relatively shorter. This was not accompanied by any differences in the ratings of the repeated tasks—for example, it did not appear that the more able *S*s found the repetition more boring, although they *said* they became easily bored.

A fourth variable is *S*'s activity or passivity during a task. Factor V seems to reflect this to some extent. The three periods in which *S* predominantly watched or listened to *E*, as opposed to working at some task, were all loaded on this factor. The *S*s who most overestimated the passive periods tended to rate them as relatively easier. It was suggested that for the more active persons, the passive periods differed more sharply from the active ones. Scores on the time illusions, performance scores, and scores on the MMPI appeared consistent with the interpretation of Factor V as reflecting an active–passive dimension, although the evidence was not entirely clear-cut. The questionnaire results seemed also to support this interpretation.

A fifth variable—time perspective—appeared in the questionnaire items related to Factor I. Individuals for whom past and future were subjectively less real, overestimated the time occupied by present activity. Since this is not an effect of the activity as such, but rather relates to individual differences in the frame of reference of time judgments, it is of secondary interest to the present discussion, which will continue mainly in terms of the first four variables.

We may consider these four variables as constituting four sources of overestimation of time: time may seem long during an interval because the activity is boring, because attention is being paid to the passage of time, because the activity is unfamiliar, or because *S* is relatively passive. Obviously many situations in which time is overestimated will involve more than one of these variables. For example, waiting for someone who is late for an appointment might involve most or all of these factors.

Are these variables adequate to account for the results of the experiments reported in the literature on moderate[13] time intervals, or need additional factors be considered?

The experimental results prior to 1930, which are summarized in Table 11, can fairly easily be handled in these terms. The active–passive distinction appears mos marked, but many of the differences presumably would involve an interesting–boring, or an attention-to-time dimension as well. One or both of the latter dimensions is required to handle Axel's (1924) results: overestimation of active tasks of a repetitious sort (tapping and writing i's).

The case of pain presents an interesting paradox. Enduring pain is something that is ordinarily thought of as making time seem long. And yet in experiments, periods of pain have either been rather slightly

TABLE 11

SUMMARY OF EARLY EXPERIMENTS ON THE ESTIMATION OF MODERATE TIME INTERVALS

Experimenter	Estimated as Relatively Long	Estimated as Relatively Short
MacDougall (1904)	listening to reading waiting in idleness	marking m's estimating time
Yerkes & Urban (1906)	idleness listening to reading estimating time	writing from dictation
Spencer (1921)	listening to prose listening to poetry reading prose aloud reading poetry aloud	writing from dictation
Sturt (1923)	looking out window reading aloud	listening to clock pain
Axel (1924)	estimating time tapping writing 1's	cancellation analogies number series adding 7's
Swift & McGeoch (1925)	idleness listening to reading	copying ns. syllables copying fiction copying catalog
Gulliksen (1927)	resting experiencing fatigue hearing slow metronome hearing fast metronome pain	mirror reading writing from dictation long division
Whitely & Anderson (1930)	empty interval listening to buzzer	listening to music

---

[13] The writer classifies as short those intervals falling within the "conscious present" (up to six seconds or so), and as moderate, those from this length up to about 10 minutes.

overestimated (Gulliksen, 1927) or actually underestimated (Sturt, 1923). The critical difference would seem to be that in these experiments the *S*s were deliberately inflicting temporary pain on themselves, thus transforming the pain from a dull, passive experience to an active, challenging one.

The only cases in Table 11 which seem to present a serious problem to the present set of variables are the two instances where reading aloud by *S* was overestimated. This seems, at least superficially, to be an activity which is interesting rather than boring, active rather than passive, and filled rather than empty. Some novelty may be involved, but it is not obvious that this would be sufficient to outweigh the other factors. Two hypotheses suggest themselves. One is that reading really does involve a strong passive element—after all, it is always the author who decides what is coming next. The other is that in literary narrative the time scale is usually condensed—events which occupy a given period of time are reported in a shorter period of reading time. If *S*'s estimates are influenced by the time scale of what he is reading, the result will tend to be "overestimation" of the objective interval.

In the more recent literature on the estimation of moderate time intervals,[14] the present set of factors seems capable of dealing with many of the findings, but there are some new issues raised.

Our factor relative to repeated tasks appears to represent the similar effect reported by Harton (1938, 1939a), Eson and Kafka (1952), and Falk and Bindra (1954). Our interesting–boring dimension readily subsumes Harton's (1938) effect of difficulty, if his difficult tasks are assumed to have been more interesting to his *S*s.

Experiments concerning stress and anxiety present more of a problem (Falk & Bindra, 1954; Harton, 1939a; Henrikson, 1948; Simmel;[15] Smith, Wing, & Jerison, 1955). Although our experimental situation was not conducive to extreme degrees of stress, the activities were intended to vary along this dimension. Some of the tasks were given as "tests" involving time pressure; others were neutral in this regard; some of the activities were designed to be pleasant and relaxing. Surely the subjects were differentially prone to anxiety under pressure,[16] but no stress factor emerged from their estimates.

In the experiments reported in the literature in which time was judged longer under stress, the effect appears in connection with repeated trials, and it is possible that it was indirect. In Falk and Bindra's (1954) experiment, for example, the occasional shock may have acted mainly by counteracting the effects (of habituation, or whatever) which led to increasing underestimation over their series of trials. Correspondingly, in Simmel's experiment[17] most of the obtained difference was due to the failure of his experimental group to show a decline in its estimates during a second period, one in which ego-involving instructions were given. The experiment of Smith et al. (1955) involved a long series of trials, and the stress imposed by a higher noise level may have worked in a fashion similar to Falk and Bindra's shock. Harton's (1939a) results were also obtained over repeated trials. In the one experiment in which repeated trials were *not* involved (Henrikson, 1948), the results were actually in the opposite direction (time was underestimated during public speaking).

There are numerous instances in everyday experience where time seems long under stress. Such instances may conceivably be explained in similar fashion; i.e., the slow passage of time in stressful situations may result from a failure of habituated modes of response in such situations. However, others of the factors may be involved as well. Boredom hardly seems a likely candidate, but an inhibition of activity by anxiety would seem to be a possibility, or an acute attention to the passage of time in the anticipation of threatening events. At any rate, the influence of stress on time perception stands as a matter in need of closer experimental analysis.

---

[14] For a more complete review, see Loehlin (1957, pp. 11–21).

[15] E. C. Simmel. The effect of stress-producing instructions on time perception in a group test situation. Personal communication, 1956.

[16] The σ of MMPI *Pt* in our sample was 15.4. Goodstein (1954) reports a median *Pt* σ of 10.3 in eight studies of college males.

[17] See Footnote 15.

Another phenomenon which has been described in the literature is the increasing underestimation of time as a goal is approached (Hindle, 1951). The present investigation was not well suited to detecting this since, for other reasons, it was desirable to make most of the tasks of indefinite length, i.e., comparable to Hindle's "no goal" condition. Consequently no safe conclusion can be drawn as to whether or not such a phenomenon could be accounted for by the present factors. Certainly one would suppose that the interest of a task and the level of activity might rise as a goal was neared, but so might attention to time. One would also need to take into account the influence of a goal on the time perspectives involved.

The results of the present investigation bear on another topic which has been the subject of some discussion in the literature —that of sex differences in the estimation of time. Several of the earlier studies (Axel, 1924; Gulliksen, 1927; MacDougall, 1904; Yerkes & Urban, 1906) reported a greater overestimation of time by women than by men. In more recent work, this difference has not, as a rule, been found (Gilliland & Humphreys, 1943; Harton, 1939b; Swift & McGeoch, 1925). The present study disclosed a factor of sex differences (Factor VI), but its loadings were mostly on performance scores, rather than time estimates. On the other factors, male and female *S*s were about equally distributed between highs and lows, as Table 12 shows. Thus the present investigation tends to agree with the other more recent studies which have failed to find the sex differences in time estimation reported in the earlier experiments. It has been suggested that changes in the role of women in the American culture have been responsible for the disappearance of this difference (Gilliland, Hofeld, & Eckstrand, 1946).

TABLE 12

SEX OF *S*S WITH HIGH AND LOW FACTOR SCORES

Group	Factor			
	I	II	IV	V
High scorers				
Men	6	6	9	6
Women	22	22	19	22
Low scorers				
Men	5	9	6	9
Women	23	19	22	19

It is possible that the sex difference reported in the early studies may still be found among children. In the Cattell and Gruen (1955) factorial study of 11-year-olds mentioned earlier, female sex was loaded on the same factor as high time estimates. However, Gilliland and Humphreys (1943) failed to find significant sex differences in time estimation among 48 fifth-grade children, so the issue remains doubtful.

Short Interval Estimates and Other Time Judgments

Let us turn now to the second main question to which the present investigation was directed: To what extent do the obtained factors apply to estimates of shorter and longer intervals, and other time judgment phenomena?

Perhaps the most striking feature of the results in this respect is the absence of a general time estimation factor. Although the first factor loaded all the 2-minute estimates and the 20-minute estimate, the short interval estimates came out on a separate factor.[18] People who overestimate minutes do not necessarily overestimate seconds. The subjective scale of time is apparently not a highly integrated dimension for most of our *S*s.

---

[18] An oblique factor solution would probably have shown a slight degree of relationship between the two factors. Their factor scores correlated .14.

Note that this does not imply that moderate and short intervals have nothing in common. Several 2-minute estimates do load on Factor II, and meaningfully so. But the influences which set the *general* level of an individual's estimates in the 2-minute periods do not appear to have acted in the same way on his estimates of the shorter intervals. In some respects, of course, this is not surprising. The degree of general interest or boredom contributed to Factor I, but it seems reasonable to suppose that variations in boredom will not be great for periods as short as one and four seconds.

Are there any data in the literature which bear on this relative independence of moderate and short interval estimates?

A general tendency toward positive intercorrelation of estimates of moderate intervals (½ minute to 3 minutes) is present in Adler's (1954) data. Clausen (1950) has reported high intercorrelations of estimates of intervals from 5 seconds to 15 seconds. (His subjects were schizophrenics, however, and some caution must be exercised in generalizing to other samples. Eccentricities of judgment—e.g., a subject who consistently estimates a period of seconds as several hours—could inflate correlations considerably in such a group.) Eson and Kafka (1952) report some correlations between a 15-second and a 2-minute interval. These *r*'s ranged from .69 to .73 under four conditions. This is the only case known to the writer in which correlations between estimates of intervals this far apart are reported—although 15 seconds is really not in the short interval class. These *r*'s are for estimates made *under the same conditions* for the 15-second and the 2-minute intervals: the case is perhaps more closely analogous to the relationship of short and moderate estimates on our Factor II than to the relationship between our Factors I and II.

Until more complete evidence is forthcoming, then, our tentative conclusion will be that the estimation of moderate and short intervals of time proceeds in substantially different ways, but that at least some content conditions affect both similarly.

The comparison and estimation of short intervals also came out on separate factors. This result is in line with earlier findings: Clausen (1950) found the reproduction method (which involves comparison) to give rather low correlations with results from two methods involving estimation. Although a negative time error in comparing time intervals has traditionally been called an "overestimation" of time, it would appear that it is at least a different kind of overestimation than that which occurs with absolute estimates.

The fact that both of the comparison procedures would have essentially all of their common variance on a single factor was somewhat unexpected. Although the general form of the stimulus presentation and judgment was the same in both cases, response tendencies such as responding at random, alternation of responses, or preference for one or the other of the two judgments, would have had quite different effects on the scores on the two variables. Both of the tasks, however, involved "illusions"—the apparent lengthening of the second interval and of the decelerating sound pattern. Hence, some aspect of *S*'s attitude which would make him more or less subject to such illusions seems a likely explanation for the similar relationships of these two measures to the other variables. In the light of the rest of the loadings on Factor V, it was suggested that this common aspect might be a relatively active or passive perceptual attitude.

Some evidence in support of such a view is offered in the work of Holzman (1954) and of Holzman and Klein (1950, 1954). Their "levelers" displayed larger time errors in visual brightness, lifted weights, and auditory loudness than their "sharpeners" did, and also displayed generally passive personality characteristics. Woodrow (1935), on the basis of his work with practice effects, has argued that time errors in time judgment are closely related to those in other modalities.

Conclusion: Method and Results

In evaluating the results of this investigation, which represents a somewhat novel approach to an old problem, some comment is necessary concerning several limitations of our study, some of which are specific to it, some inherent in this use of the factor-analytic method, and some, general methodological limitations common to most psychological research today.

1. Neither the situations nor the persons represent truly random samples from clearly defined populations, making generalizations to some extent hazardous. These difficulties exist, of course, in most psychological research (Brunswik, 1949). Factor analysis, to the extent that it usually employs a greater variety of situational variables than most techniques, is perhaps less subject to this stricture than they. As for the present study, it at least used an array of situations considerably larger than that of any previous investigation in this area known to the writer.

2. Correlations depend on individual differences. As mentioned earlier, a situational variable which tends to be uniform in its effects may have a relatively slight influence on the correlations while having an important effect on the estimates themselves. Something of the kind may have happened in the present study, with the variable "interest vs. boredom," which was markedly related to differences between mean estimates,[19] but did not appear to carry as much weight in determining the factors as might have been expected.

3. As discussed in an earlier section, various sources of common variance somewhat irrelevant to the main purposes of this study played a significant role in the factor resolution. While this must be regarded as a methodological fault, it does not seem to have prevented the obtaining of meaningful factors.

4. The interpretation of factors has at best its subjective and intuitive aspects—as does the interpretation of the results of most experiments, if more than trivial conclusions are drawn. In the present investigation, provision was made for obtaining additional evidence—*S*s' ratings of the tasks and MMPI scores—to reduce the risks inherent in this procedure. Also, the relationships of the factors to the items from the time questionnaire provided additional information concerning the adequacy of these interpretations.

On the credit side it appears that, despite the above limitations, the investigation proved fairly successful. Variables were isolated which seemed to relate meaningfully both to the results of earlier investigations and to our *S*s' statements about their behavior and perception relevant to time in everyday situations. Some light was shed on methodological questions: the results yielded some information concerning the extent to which, and the conditions under which, results obtained with different methods of measurement, or involving different time intervals, can be expected to reflect the same underlying variables. Finally, at least scattered information was obtained relevant to the integration of time perception into the larger system of personality.

For the reasons discussed, it cannot be claimed that the set of variables isolated in this study represents a definitive description of the variables affecting time perception, yet the present study appears to represent a step in that direction, and one taken from a broader empirical base than experimental investigations in this area have used in the past.

SUMMARY

Previous investigators have shown that estimates of time during periods occupied with different activities will be influenced by the content of the intervals, but little agreement exists as to the important variables involved.

This investigation sought to isolate such variables by a factor analysis of time esti-

---

[19] The Spearman rank-order correlation between mean rating of "boring" and mean time estimate, over the sixteen 2-minute activities, was +.61. An analysis of the relationships between estimates and ratings may be found in Loehlin (1957, pp. 73–78).

mates made under a number of conditions selected to be representative of those used by earlier writers.

One hundred and five college students made estimates of the apparent duration of sixteen 2-minute activities. In addition, they (*a*) rated each activity on several dimensions, (*b*) made estimates of shorter and longer intervals, (*c*) made judgments of short time intervals by comparison methods, and (*d*) completed a questionnaire concerning their perception of and attitudes toward time in everyday situations. Seventy-four of them also completed the MMPI.

The 16 time estimates plus 12 additional measures were intercorrelated and factored by the principal axes method. Seven of eight rotated factors were interpreted. The main variables contributing to the apparent length of the time intervals appeared to be:

1. interest vs. boredom
2. filled vs. empty intervals (attention to time?)
3. repetition of an activity
4. activity vs. passivity

A factor of stress or anxiety was suggested by previous work, but not found. A factor of sex differences was obtained, but little of its variance was on time estimates.

Evidence from the ratings of the activities and the MMPI scores was used in interpreting the factors; evidence from the time perception questionnaire in general supported and supplemented the interpretations.

REFERENCES

ADLER, N. Perception of time as a function of self-organization. Unpublished doctoral dissertation, Univer. of California, Berkeley, 1954.

AXEL, R. Estimation of time. *Arch. Psychol.,* 1924, **12**, 1–72.

BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.,* 1953, **17**, 327–333.

BECKMAN, F. S., & QUARLES, D. A., JR. Multiple regression and correlation analysis on the IBM type 701 and type 704 electronic data processing machines. *Amer. Statistician,* 1956, **10**, 6–8, 16.

BONAPARTE, MARIE. Time and the unconscious. *Int. J. Psycho-Anal.,* 1940, **21**, 427–468.

BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of California Press, 1949.

CATTELL, R. B. *Factor analysis.* New York: Harper, 1952.

CATTELL, R. B., & GRUEN, W. The primary personality factors in 11-year-old children, by objective tests. *J. Pers.,* 1955, **23**, 460–478.

CLAUSEN, J. An evaluation of experimental methods of time judgment. *J. exp. Psychol.,* 1950, **40**, 756–761.

DOOLEY, LUCILE. The concept of time in defence of ego integrity. *Psychiatry,* 1941, **4**, 13–23.

DUBOIS, F. S. The sense of time and its relation to psychiatric illness. *Amer. J. Psychiat.,* 1954, **111**, 46–51.

ESON, M. E., & KAFKA, J. S. Diagnostic implications of a study in time perception. *J. gen. Psychol.,* 1952, **46**, 169–183.

FALK, J. L., & BINDRA, D. Judgment of time as a function of serial position and stress. *J. exp. Psychol.,* 1954, **47**, 279–282.

GILLILAND, A. R., HOFELD, J., & ECKSTRAND, G. Studies in time perception. *Psychol. Bull.,* 1946, **43**, 162–176.

GILLILAND, A. R., & HUMPHREYS, DOROTHY WINDES. Age, sex, method, and interval as variables in time estimation. *J. genet. Psychol.,* 1943, **63**, 123–130.

GOODSTEIN, L. D. Regional differences in MMPI responses among male college students. *J. consult. Psychol.,* 1954, **18**, 437–441.

GULLIKSEN, H. The influence of occupation upon the perception of time. *J. exp. Psychol.,* 1927, **10**, 52–59.

GUTTMAN, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954. (a)

GUTTMAN, L. Some necessary conditions for common-factor analysis. *Psychometrika,* 1954, **19**, 149–161. (b)

HARTON, J. J. The influence of the difficulty of activity on the estimation of time. I and II. *J. exp. Psychol.,* 1938, **23**, 270–287; 428–433.

HARTON, J. J. The influence of the degree of unity of organization on the estimation of time. *J. gen. Psychol.,* 1939, **21**, 25–49. (a)

HARTON, J. J. The relation of time estimates to the actual time. *J. gen. Psychol.,* 1939, **21**, 219–224. (b)

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory* (rev. ed.), and *Supplementary Manual.* New York: Psychological Corp., 1946.

HENRIKSON, E. H. A study of stage fright and the judgment of speaking time. *J. appl. Psychol.,* 1948, **32**, 532–536.

HINDLE, HELEN M. Time estimates as a function of distance traveled and relative clarity of a goal. *J. Pers.,* 1951, **19**, 483–501.

HOLZMAN, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. *J. Pers.,* 1954, **22**, 375–394.

HOLZMAN, P. S., & KLEIN, G. S. The "schematizing process": Perceptual attitudes and personality qualities in sensitivity to change. *Amer. Psychologist,* 1950, **5**, 312. (Abstract.)

HOLZMAN, P. S., & KLEIN, G. S. Cognitive system-principles of leveling and sharpening: Individual differences in assimilation effects in visual time-error. *J. Psychol.,* 1954, **37**, 105–122.

LOEHLIN, J. C. The influence of different activities on the apparent length of time. Unpublished doctoral dissertation, Univer. of California, Berkeley, 1957.

MACDOUGALL, R. Sex differences in the sense of time. *Science,* 1904, **19**, 707–708.

ROSENZWEIG, S., & KOHT, AASE GRUDE. The experience of duration as affected by need-tension. *J. exp. Psychol.,* 1933, **16**, 745–774.

SCHILDER, P. Psychopathology of time. *J. nerv. ment. Dis.,* 1936, **83**, 530–546.

SMITH, A. K., WING, S., & JERISON, H. T. Effect of acoustic noise on time judgment. *Amer. Psychologist,* 1955, **10**, 428–429. (Abstract.)

SPENCER, L. T. An experiment in time estimation using different interpolations. *Amer. J. Psychol.,* 1921, **32**, 557–562.

STURT, MARY. Experiments on the estimate of duration. *Brit. J. Psychol.,* 1923, **13**, 382–388.

SWIFT, E. J., & MCGEOCH, J. A. An experimental study of the perception of filled and empty time. *J. exp. Psychol.,* 1925, **8**, 240–249.

THOMSON, G. H. *The factorial analysis of human ability.* (5th ed.) London: Univer. of London Press, 1951.

THURSTONE, L. L. *The vectors of mind.* Chicago: Univer. of Chicago Press, 1935.

WHITELY, P. L., & ANDERSON, J. C. The influence of two different interpolations upon time estimation. *J. gen. Psychol.,* 1930, **4,** 391–401.

WOODROW, H. The effect of practice upon time-order errors in the comparison of temporal intervals. *Psychol. Rev.,* 1935, **42,** 127–152.

WRIGLEY, C. F. An empirical comparison of various methods for the estimation of communalities. Contract Report No. 1, Contract AF 41 (657)–76, Univer. of California, Berkeley, 1956. (*Educ. psychol. Measmt.,* in press.)

YERKES, R. M., & URBAN, F. M. Time-estimation in its relation to sex, age, and physiological rhythms. *Harvard psychol. Stud.,* 1906, **2,** 405–430.

(Accepted for publication April 29, 1958)

APPENDIX A

Means and Standard Deviations of the Variables

No.	Variable	Units	Mean	*SD*
1	2-min. est., Introduction	Secs/10	15.6	6.4
2	2-min. est., Writing 1's (1st)	Secs/10	19.8	7.1
3	2-min. est., Easy anagram	Secs/10	16.2	7.1
4	2-min. est., Difficult comparisons	Secs/10	17.2	7.3
5	2-min. est., Gottschaldt practice	Secs/10	16.4	8.2
6	2-min. est., Gottschaldt test	Secs/10	15.1	6.1
7	2-min. est., Slow calculation	Secs/10	16.8	7.7
8	2-min. est., Listening to reading	Secs/10	15.7	7.2
9	2-min. est., Pleasant thoughts	Secs/10	20.1	7.8
10	2-min. est., Interruption	Secs/10	14.3	6.2
11	2-min. est., Writing 1's (2nd)	Secs/10	20.5	7.0
12	2-min. est., Easy comparisons	Secs/10	15.6	6.7
13	2-min. est., Hard anagram	Secs/10	13.9	5.7
14	2-min. est., Counting "the's"	Secs/10	15.3	7.1
15	2-min. est., Concentration on time	Secs/10	23.6	9.9
16	2-min. est., Questionnaire	Secs/10	13.1	6.7
17	Estimate of 1 sec.	Sec. × 10	6.9	4.7
18	Estimate of 4 sec.	Sec. × 10	39.7	19.5
19	Estimate of 20 min.	Min.	15.9	6.6
20	Neg. time error, Diff. comp.	Responses "2"	9.1	2.4
21	Illusion, Accel.-decel.	Trials Dec. > Acc.	8.6	1.9
22	1's written in 4 min.	No./10	61.9	15.8
23	Questionnaire items in 2 min.		25.0	5.6
24	Easy anagram solutions		16.5	5.0
25	Gottschaldt figures solutions		9.1	4.6
26	Error, counting "the's"		8.1	4.7
27	Persons in *S*'s group		5.5	2.1
28	Age of *S*	Yr.	19.4	1.1

APPENDIX B

INTERCORRELATION MATRIX

Var.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Communality Estimate
1		17	30	16	25	32	18	18	16	32	26	15	15	18	08	15	00	06	29	−16	−13	−01	21	−12	07	−12	02	−03	38
2	17		41	36	30	38	19	24	19	16	34	15	16	06	22	25	12	28	25	−06	−11	−11	−10	−02	01	01	11	−03	50
3	30	41		36	47	47	20	32	35	32	32	42	44	26	18	35	17	11	22	−10	−06	12	−02	−13	−05	02	07	15	56
4	16	36	36		47	44	56	33	30	30	37	46	38	45	39	44	05	12	30	00	00	07	−05	−09	−12	−03	02	−03	55
5	25	30	47	47		60	34	18	28	25	33	31	38	40	39	27	09	15	28	05	−11	−05	−09	−22	−16	03	00	13	62
6	32	38	47	44	60		35	22	31	25	25	29	34	33	25	32	−02	00	33	−09	−10	15	18	−06	−03	−05	05	05	59
7	18	19	20	56	34	35		47	36	27	23	44	38	49	30	42	−05	01	03	03	01	12	−06	−10	05	−08	−05	00	57
8	18	24	32	33	18	22	47		32	29	22	32	36	42	20	32	12	16	13	−04	−17	12	−13	−24	08	01	15	08	45
9	16	19	35	30	28	31	36	32		38	31	40	46	36	49	31	05	11	−03	−02	05	14	−26	−29	12	−12	03	16	53
10	32	16	32	30	25	25	27	29	38		23	39	45	38	30	38	01	10	06	−09	−05	18	−14	−18	18	−10	07	−03	46
11	26	34	32	37	33	25	23	22	31	23		54	56	38	29	26	20	28	22	02	03	−02	−03	−10	06	−07	−04	01	59
12	15	15	42	46	31	29	44	32	40	39	54		53	57	38	41	10	11	18	06	11	−02	−07	−20	−04	−12	−02	10	62
13	15	16	44	38	38	34	38	36	46	45	56	53		39	33	31	01	07	−03	09	−04	15	−11	−01	15	−14	−07	07	65
14	18	06	26	45	40	33	49	42	36	38	38	57	39		47	53	05	06	16	01	00	05	−12	−36	12	−08	07	−04	66
15	08	22	18	39	39	25	30	20	49	30	29	38	33	47		36	18	37	−04	06	08	−16	−17	−33	01	−01	−03	−14	58
16	15	25	35	44	27	32	42	32	31	38	26	41	31	53	36		02	13	10	−01	01	05	01	−14	−06	11	09	05	52
17	00	12	17	05	09	−02	−05	12	05	01	20	10	01	05	18	02		62	00	01	−08	−16	−06	−25	−04	08	−10	−07	49
18	06	28	11	12	15	00	01	16	11	10	28	11	07	06	37	13	62		−01	03	−06	−17	00	−20	00	−07	−07	−04	57
19	29	25	22	30	28	33	03	13	−03	06	22	18	−03	16	−04	10	00	−01		−16	−17	−07	00	−13	−19	−08	14	10	43
20	−16	−06	−10	00	05	−09	03	−04	−02	−09	02	06	09	01	06	−01	01	03	−16		22	−23	−12	−08	12	19	−07	09	24
21	−13	−11	−06	00	−11	−10	01	−17	05	−05	03	11	−04	00	08	01	−08	−06	−17	22		−13	−01	−14	−06	04	−10	−03	24
22	−01	−11	12	07	−05	15	12	12	14	18	−02	−02	15	05	−16	05	−16	−17	−07	−23	−13		08	19	−04	−05	12	−08	34
23	21	−10	−02	−05	−09	18	−06	−13	−26	−14	−03	−07	−11	−12	−17	01	−06	00	00	−12	−01	08		30	00	−17	04	−11	39
24	−12	−02	−13	−09	−22	−06	−10	−24	−29	−18	−10	−20	−01	−36	−33	−14	−25	−20	−13	−08	−14	19	30		−08	−17	−02	−15	46
25	07	01	−05	−12	−16	−03	05	08	12	18	06	−04	15	12	01	−06	−04	00	−19	12	−06	−04	00	−08		−11	00	05	31
26	−12	01	02	−03	03	−05	−08	01	−12	−10	−07	−12	−14	−08	−01	11	08	−07	−08	19	04	−05	−17	−17	−11		−05	−11	31
27	02	11	07	02	00	05	−05	15	03	07	−04	−02	−07	07	−03	09	−10	−07	14	−07	−10	12	04	−02	00	−05		−02	14
28	−03	−03	15	−03	13	05	00	08	16	−03	01	10	07	−04	−14	05	−07	−04	10	09	−03	−08	−11	−15	05	−11	−02		29

APPENDIX C

UNROTATED FACTOR LOADINGS

Variable	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	35	−25	20	14	13	−04	29	16	−03	15	03	04
2	43	−05	40	01	01	−02	−24	−01	−31	−01	24	−10
3	60	−17	21	02	10	−20	−20	−01	17	18	−00	−14
4	67	−11	04	−17	−20	17	−06	−10	−08	−12	02	00
5	63	−07	24	−28	−08	−14	−06	22	04	−05	−18	18
6	59	−38	20	−13	−08	−04	−04	26	−02	04	−12	−01
7	60	−10	−27	−08	−11	22	−04	−08	−22	−08	−21	−02
8	53	−00	−08	06	29	15	−08	−24	−15	04	−14	01
9	59	12	−24	08	18	−13	−12	18	06	−18	05	−14
10	54	−06	−19	20	21	03	02	12	07	09	15	04
11	59	09	12	22	−24	−22	10	−19	01	04	17	14
12	69	07	−17	02	−19	−11	16	−24	16	−00	04	−08
13	66	−04	−27	27	−15	−27	−16	04	03	04	−02	15
14	69	07	−26	−10	04	20	29	−05	04	03	−04	11
15	57	39	−04	−02	−10	18	01	31	01	−16	11	−01
16	59	−02	−09	−11	−02	25	−03	−06	02	20	05	−21
17	16	48	42	23	04	09	−05	−10	14	05	−18	03
18	26	49	44	33	−02	14	−02	−01	−02	−06	−10	−08
19	27	−28	40	−25	12	−08	26	−18	00	−11	11	07
20	−02	31	−16	−18	−20	−18	−04	−00	−18	18	−08	02
21	−05	23	−21	−16	−26	−07	08	05	08	07	12	−21
22	06	−39	−22	18	12	16	−23	−02	23	−03	−05	05
23	−13	−36	18	23	−27	14	23	10	02	18	−15	−16
24	−31	−44	−01	26	−35	04	−21	−06	−08	−05	03	00
25	03	08	−24	29	18	−13	12	12	−32	15	−02	04
26	−08	20	04	−33	04	11	−25	−02	05	34	06	14
27	05	−17	03	−03	23	11	02	−06	−04	00	14	−08
28	07	−01	−04	−13	18	−42	02	−11	−04	−07	−22	−22

APPENDIX D

ROTATING MATRIX

Factor	1	2	3	4	5	6	7	8
1	86	19	−36	−09	16	−10	−18	13
2	14	69	58	36	−05	20	−04	−01
3	33	−40	61	−36	−33	−01	08	34
4	18	−22	−05	69	−22	−45	43	09
5	15	−45	26	27	73	28	−04	−12
6	06	−18	−28	31	−41	75	−10	21
7	−03	−22	12	24	−25	−27	−84	−21
8	25	−08	03	−11	−28	16	22	−88

APPENDIX E

ROTATED FACTOR LOADINGS

Variable	I	II	III	IV	V	VI	VII	VIII
1	16	02	32	33	31	−03	−20	−02
2	26	27	38	−02	21	−07	26	07
3	48	06	34	13	25	−16	25	07
4	50	03	38	−04	03	19	04	35
5	51	05	58	00	−01	−12	02	−01
6	41	−12	63	13	18	07	07	−02
7	56	−12	11	−01	04	30	−03	29
8	49	07	−07	−05	36	02	00	26
9	70	02	−04	13	07	−02	00	−13
10	55	−01	−01	22	26	12	−08	−04
11	40	26	18	41	−08	−06	10	34
12	59	04	08	27	−11	01	−06	44
13	67	−04	02	42	−06	09	26	06
14	64	−02	07	07	06	16	−39	30
15	58	36	15	−02	−17	21	−21	−10
16	52	03	17	−09	11	22	−08	25
17	05	70	−02	−02	05	−09	02	07
18	11	78	04	08	05	02	00	04
19	01	−06	46	02	28	−30	−15	30
20	11	03	−10	−08	−43	−12	00	00
21	05	−03	−07	−03	−43	02	−11	02
22	08	−29	−09	03	32	28	22	−02
23	−35	−05	27	31	07	29	−02	04
24	−40	−22	05	18	−01	33	43	04
25	14	−02	−28	28	06	−01	−11	−19
26	02	06	00	−46	−11	−07	06	−03
27	02	−09	02	−07	29	−01	−06	04
28	14	−16	−04	08	00	−44	04	02

Vol. 73, No. 5 Whole No. 475, 1959

Psychological Monographs: General and Applied

THE ONTOGENY OF CERTAIN LOGICAL OPERATIONS:

PIAGET'S FORMULATION EXAMINED BY NONVERBAL METHODS[1]

MARTIN D. S. BRAINE[2]

New York University

THE EXPERIMENTS reported are concerned with the work of Piaget on the development of intelligence in children. Piaget's theory has two aspects. First, it is a theory of intelligence: Piaget believes that the development of intelligence consists in the development of an ability to perform logical operations. Second, Piaget makes specific statements about the ages at which certain types of reasoning develop. He claims that as children grow older certain specified groups of logical operations develop in the average child at given ages. In his studies the principal ages of transition are reported to be at 7 and 11 yr., approximately. At around age 7, the operations of the spatial and class interpretations of the Boolean calculus make their appearance in children's thinking; at around age 11, the operations of the propositional interpretation appear.[3] The appearance of the operations of the spatial interpretation are revealed in Piaget's studies by changes at around age 7 yr. in the child's manner of using measuring instruments and in his performance on a number of tasks supposedly demanding conception of, or reasoning about, spatial relations (Piaget & Inhelder, 1956; Piaget, Inhelder, & Szeminska, 1948). The operations of the class interpretation are revealed by the development of number concepts and by the disappearance of certain anomalies in the child's class concepts (Piaget & Szeminska, 1952). The appearance of propositional operations at age 11 is attested, according to Piaget, by the child's performance on a variety of reasoning tasks (Piaget, 1955; Piaget & Inhelder, 1953).

The notion that the nature of intelligence lies in the performance of logical operations is not new in psychology. It is clearly implicit in Spearman's (1923) identification of intelligence with "eduction of relations" and "eduction of correlates."[4] Moore (1929) seems to have had a similar view. Attempts to find logical models for schizophrenic thinking (Arieti, 1948; Von Domarus, 1944) also implicitly make the assumption that the only proper model for "normal" reasoning was described by Aristotle.

---

[1] This study is based upon a doctoral dissertation submitted to the Department of Psychology, New York University Graduate School of Arts and Science. The writer wishes to express his thanks to Elsa Robinson who sponsored the research, and to H. L. Teuber for his invaluable suggestions and criticisms. The writer is also indebted to Marjorie Cornell and Elsie Stange for making available the subjects and the facilities of their nursery schools. The work was partially supported by the Commonwealth Fund of New York through their grant to H. L. Teuber and the Psychophysiological Laboratory, New York University-Bellevue Medical Center.

[2] Now at State University of New York.

[3] For the calculus and its various interpretations, see, for example, Langer (1953). Piaget's logical system differs from the Boolean calculus, but the differences can be neglected for the purposes of this study. The interested reader may refer to his treatise on logic (Piaget, 1949).

The operations of the Boolean calculus are called *conjunction* and *disjunction.* A conjunction of classes is denoted in the English language by the qualification of a noun by an adjective, as, for instance, "white sheep" (i.e., objects that are *both* white, *and also* sheep). A disjunction of classes is often expressed by the words *and* and *or,* as, for instance, "men *and* women," "barren mountains, *or* fertile valleys."

A spatial conjunction is expressed in the two following examples: "The United Nations building is located at 42nd Street *and* the East River Drive," "a straight line is the region *common to* two intersecting planes." Spatial disjunction indicates partition or composition, e.g., "Canada *and* the United States make up North America."

A conjunction of propositions asserts the joint truth of two propositions, often by the use of the word *and,* for example, "Fabius was born at the rising of the dog-star, *and* Fabius will perish in the sea." A disjunction of propositions carries the meaning "either . . . or . . . (or both)."

Piaget's theory differs from those of previous investigators, first, in that it aims to give a description of the *development* of intelligence, and second, in that his logical model for reasoning is worked out in great detail and is derived from modern symbolic logic (Piaget, 1949). Piaget attempts to support his theory with a good deal of evidence, although the validity of some of it has been questioned (Fraisse & Vautrey: 1952a, 1952b).

Testing Piaget's theory in detail will obviously prove a monumental task, not only because of its tremendous scope, but also because some of his concepts do not have a sufficiently firm basis in experimental operations. The investigation reported here is intended as an attempt to examine the validity of Piaget's theory within a limited context. His views on the development of length measurement and of concepts of order in children have been singled out for study. The development of length measurement is discussed in the next section and order concepts in the section following. In both these sections special attention is paid to Piaget's views on the specific ages at which particular types of responses develop.

The ontogenesis of measurement provides a particularly convenient point of departure since the logic of measurement has been well analyzed by workers interested in the bases of science (Bergmann & Spence, 1944; Campbell, 1920; Nagel, 1930). In the next section an experiment is reported in which the development of a particular inferential response in children is traced. This response is the inference that a stick, A, is longer than another stick, C, when it has been demonstrated that A is longer than a measuring rod, B, and that B is in turn longer than C. In logical symbolism this may be written $A > B . B > C . \supset . A > C$.[5] According to Nagel and other logicians this inference is one of those fundamental in length measurement (i.e., all length measurement tacitly assumes the validity of this inference).

In the section entitled The Development of a Concept of Position Order, the development of discrimination of position order in children is studied. An experiment is reported in which *S* is invited to match a sample card on which colored discs have been pasted in a given order, say, red-gray-green, with one of three other cards in which the order of the colors is gray-red-green, red-green-gray, and red-gray-green.

In the section which follows, the relationship between the development of the inferential response $A > B . B > C . \supset . A > C$ and of the discrimination of position order is investigated. Piaget believes that the development of inferences such as $A > B . B > C . \supset . A > C$ is one aspect of the development of a more general ability to perform logical operations. The task Piaget has set himself is to give a detailed description of these logical operations and of their formal interrelations. He believes that this will enable us "to construct by means of the algebra of logic a deductive theory to explain some of the experimental findings of psychology" (Piaget, 1953b, p. 26). Although Piaget does not make this explicit, a basic postulate of his theory of intelligence certainly appears to be that operations, inferences, or concepts which are logically interdependent (i.e., the definition of the one involves reference to the other, or both derive from the same interpretation of a logical calculus), will be found to develop in association with each other in children's thinking. It may be noted that experimental refutation of Piaget's views about the specific ages at which particular groups of operations develop does not refute this postulate. With this postulate it may be predicted that order concepts and the inference $A > B . B > C . \supset . A > C$ develop in association in children's thinking.

---

[4] Piaget (1953b) presents a derivation of these concepts from his own theory.

[5] To be read: "*If* (A is longer than B) *and* (B is longer than C) *then* (A is longer than C).

If the postulate that logically interdependent operations develop in association is generally valid, it follows that there should be periods of transition in the development of intelligence, during which particular groups of operations develop. If such ages of transition exist—and they may not be at 7 and 11 yr. as Piaget claims—then changes in the nature of intelligence during childhood should be observable and reflected in a changing factorial composition of intelligence tests. Failure to find such changes would cast serious doubt on any theory which seeks to relate intellectual development to the performance of logical operations. In the section on ages of transition, this study concludes with an examination of the evidence for such changes reported in the literature.

THE DEVELOPMENT OF LENGTH MEASUREMENT IN CHILDREN

The part played by the logic of measurement in Piaget's description of length-measuring behavior in children is considered. Piaget's experiments are discussed and certain criticisms of his techniques are suggested. A study which uses different techniques to study the development of certain of the inferential processes involved in measurement is then reported. The results are compared with Piaget's.

Theoretical Considerations

According to many logicians, measurement is possible in a dimension only when relations between objects in the dimension fulfill certain axioms, the so-called axioms of measurement (Bergmann & Spence, 1944; Nagel, 1930). From these, certain laws may be derived. For example, the truth of the following seven laws is always tacitly assumed whenever length is measured.[6]

$$\begin{gathered} A = B . B = C . \supset . A = C \\ A = B . B > C . \supset . A > C \\ A = B . B < C . \supset . A < C \\ A > B . B = C . \supset . A > C \\ A > B . B > C . \supset . A > C \\ A < B . B = C . \supset . A < C \\ A < B . B < C . \supset . A < C \end{gathered}$$

When length is measured, it is also assumed that lengths are additive, i.e., any length is composed of shorter lengths added together. Consequently, some arbitrary unit (e.g., in., cm.) may be defined and the length of a measured object is the number of units which have to be added together to compose another object equal in length to the first. It is the existence of this additive operation that, among other things, enables one to talk meaningfully of the length of a path which is not straight, e.g., the perimeter of a triangle whose length is the sum of the lengths of the sides.

Piaget's description of the development of measurement in children consists of an account of the development of children's understanding and use of these laws.

Piaget treats these laws not only as empirical statements which are demonstrably valid, but as behavioral processes initiated by the child *S*. For example, the expression $(A + B)$ can be regarded either as a length, or as a record of the behavior of some *S*: "*S* picks up B and places one end of B next to one end of A." If *S* does not displace B physically, but does so only implicitly or "in his imagination," then the expression $(A + B)$ can be regarded as representing an implicit displacement or operation performed by *S* on A and B.

In most studies of length judgment in children, the *S*s have been asked to compare two lengths. The method used by *S* to arrive at a judgment is likely to depend both on the stimulus conditions and on the manner in which *S* is obliged to communicate his judgment to *E*.

The importance of the stimulus conditions in length judgment may be illustrated by two prototypic examples. In the first example, a length A in one room is to be compared with another somewhat shorter length C in another room by the following procedure: in the first room a convenient third object, B, is juxtaposed with A and it is found that A is longer than B; B is then carried into the other room and juxtaposed with C and it is found that B is longer than C; it is concluded that A is longer than C. For the second example, suppose that the stimulus conditions are altered so that A and C are in the same room standing upright next to each other on the top of a table; it is then immediately clear that A extends upwards beyond the highest point of C, i.e., that A is longer than C.

Let us say that in the first case A is "inferred" to be longer than C, and in the second case that A is "perceived" to be longer than C. This appears to be not inconsistent with conventional usage of these words in the English language. In the remainder of this section the word *inference* is used to refer to the kind of method by which the length judgment is arrived at in the first example, and the judgment itself is called an *inferential response*.

---

[6] A, B, and C represent any three lengths and ">," "<," and "=" designate the relations "longer than," "shorter than," and "equal in length to," respectively.

The terms *perception* and *discrimination* will often be used to designate the length judgment in the second example.

Critique of Piaget's Techniques

In Piaget's work the stimulus conditions are rarely identical with those of the two prototypic examples above; distinguishing between an inferential and a discriminatory response therefore becomes a diagnostic problem. In his experiments the stimuli to be compared have sometimes been solid objects and sometimes the paths traced by moving objects. The stimuli have been straight or bent to form a "zig-zag," and the spatial relations between the stimuli have been varied. The response has sometimes been a judgment expressed verbally, and sometimes *S* has been asked to construct a length equal to a given length, with or without uncalibrated measuring rods available. In all experiments either *E* or *S* has used words such as *big, little, short, long, same length, measure,* etc., so that a certain vocabulary development is presupposed.

The standardization of the Binet test (see Terman & Merrill, 1937) indicates that average children between the ages of 3–0 and 3–6 are able to obey a verbal instruction to pick out the longer of two sticks 2 in. and 2½ in. in length at least 82% of the time, when the sticks are placed parallel so that the longer overlaps the shorter at both ends. It is an open question whether the word *longer* is understood at this age to apply to other spatial relations between the sticks.

In two experiments by Piaget (Piaget, 1946, Ch. 3; Piaget et al., 1948, Ch. 6), the *S*s were aged 4–8 yr. The *E* moved a bead a certain distance along a wire and *S* was required to move another bead the same distance along a second wire.[7] The shape and relative positions of the wires together with the typical errors of the youngest *S*s are shown in Fig. 1. The errors made by most of the *S*s aged 4–6 yr. were always of the same nature: they moved their bead to a position, P′, directly opposite the position, P, of *E*'s bead, thus failing to take account of the differing points of departure of the two beads. The responses of these *S*s were thus correct only in arrangement "B" in Fig. 1. The *S*s of intermediate age were correct in arrangements "A" and "C," but not in "D" or "E." The *S*s aged 7 and 8 yr. were correct in all arrangements. Following their response all *S*s were given an unmarked ruler to measure the distances. The youngest *S*s just laid the ruler across the two beads perpendicular to the wires. Not until after age 7 yr. were the *S*s able to measure adequately. Piaget interpreted his findings to indicate that his youngest *S*s identified "length" with "final position occupied"; the typical errors indicated the "finalism" which is characteristic of children of this age, and which reflects the fact that the logical operations basic to measurement are not yet available to them.

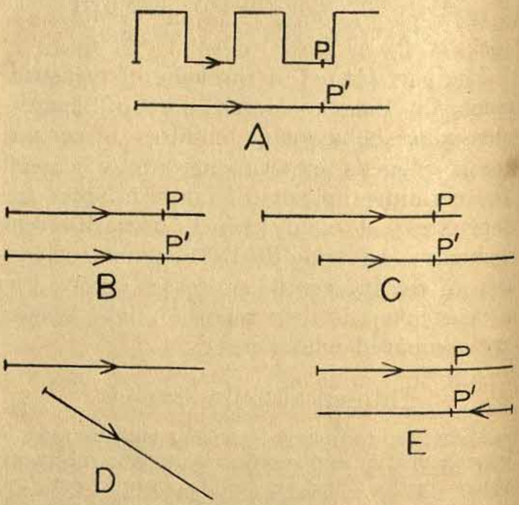


FIG. 1. Arrangements of wires used by Piaget. The arrow indicates direction of travel and P and P′ the points of arrival of the beads moved, respectively, by *E* and by the youngest group of *S*s (with whom Arrangement "D" was not used).

Fraisse and Vautrey (1952a) criticize Piaget's procedure and, in a study which also employed the paths traced by moving objects as stimuli, they found no evidence for "finalism" in young children's perception of length. Their *S*s were aged 4–6 to 5–6, the age of Piaget's youngest *S*s. Fraisse and Vautrey found that *S*s who were told to indicate the starting positions of the objects responded correctly more often than *S*s whose response was not prefaced by this instruction. Moreover, *S*s who made errors had usually indicated the starting position incorrectly, their response being appropriate to the starting position indicated. The authors suggest that the errors typical of Piaget's youngest *S*s reflect not a peculiarity of their conception of length, but their failure to notice or to recall the starting positions of the objects.

---

[7] "Veux-tu faire marcher l'autre tram pour qu'il fasse juste le même long chemin?" (Piaget, 1946, p. 56).

If the errors made by Piaget's *S*s do not constitute evidence to support his diagnosis of their length judgments as "pre-logical," then the only support for this diagnosis lies in the inappropriateness of the measuring behavior of his *S*s.

The inception of measurement is examined in greater detail in another of Piaget's studies (Piaget, 1953a; Piaget et al., 1948, Ch. 2). Piaget invited children to build with blocks a tower equal in height to a tower already built by *E*. However, the table on which *E*'s tower stood was higher than the table on which *S* was to build his tower and was some distance from it. Sticks longer, shorter, and equal to the height of the model were available to *S*s. Piaget summarizes his results as follows:

> The youngest children build up the second tower to the same visual level as the first, without worrying about the difference in height of the tables. They compare the towers by stepping back and sighting them. At a slightly more advanced stage a child lays a long rod across the tops of the two towers to make sure that they are level. Somewhat later he notices that the base of his tower is not at the same level as the model's. He then wants to place his tower next to the model on the same table to compare them. Reminded that the rules of the game forbid him to move his tower, he begins to look around for a measuring standard. Interestingly enough, the first that comes to his mind is his own body. He puts one hand on top of his tower and the other at its base, and then, trying to keep his hands the same distance apart, he moves over to the other tower to compare it. Children of about the age of six often carry out this work in a most assured manner, as if their hands could not change position on the way! Soon they discover that the method is not reliable, and then they resort to reference points on the body. The child will line up his shoulder with the top of his tower, mark the spot opposite the base on his thigh with his hand and walk over to the model to see whether the distance is the same.
>
> Eventually the idea of an independent measuring tool occurs to the child. His first attempt in this direction is likely to be the building of a third tower next to and the same height as the one he has already erected. Having built it, he moves it over to the first table and matches it against the model; this is allowed by the rules. The child's arrival at this stage presupposes a process of logical reasoning. If we call the model tower A, the second tower C, and the movable tower B, the child has reasoned that B = C and B = A, therefore A = C.
>
> Later the child replaces the third tower with a rod, but at first the rod must be just the same length as the height of the tower to be measured. He then conceives the idea of using a longer rod and marking the tower height on it with his finger. Finally, and this is the beginning of true measurement, he realizes that he can use a shorter rod and measure the height of the tower by applying the rod a certain number of times up the side.
>
> The last discovery involves two new operations of logic. The first is the process of division which permits the child to conceive that the whole is composed of a number of parts added together. The second . . . enables him to apply one part upon others and thus to build a system of units (Piaget, 1953a, pp. 77–78).

It can be seen that Piaget's diagnosis of a child's ability to perform the inferences basic to measurement is based entirely on his measuring behavior. However, it seems to the writer that a child's measuring behavior may reflect his knowledge of measuring technique as much as it does his ability to reason logically about length. Piaget's 5- and 6-yr.-old *S*s tended to use their own bodies as measuring rods. The human body is, of course, a clumsy and inadequate measuring instrument, yet the fact that a young child does not immediately perceive its imperfections and select a better tool may not necessarily indicate a deficiency in his logic. One may recall that in the cultural history of measurement it required a considerable period of time before adult human beings came to adopt more objective units of measurement than parts of the human body. Consider the behavior of the child who "measures" the height of a tower by placing one hand at its base and the other at its summit. Piaget claims that such behavior indicates deficient understanding of the logic of measurement, because the child fails to realize that it is impossible to keep the distance between the hands constant as he walks from one tower to the other: the child's measuring rod is elastic. However, the opposite point of view can be argued. To the writer such behavior suggests that the child has a rather sound knowledge of the logic of measurement, but is woefully inexpert in its practice! Such behavior appears to show a good knowledge of the seven paradigmatic rules of inference involved in measurement enumerated earlier.

Another factor which may be important in Piaget's studies of measurement is the child's motivation. Piaget relies on the inherent interest of the tasks for the child. However, it is possible that the older children are more interested than the younger *S*s in measurement and the additional accuracy it makes possible.

In another experiment (Piaget et al., 1948, Ch. 5) designed to study children's understanding of the additive operation involved in length measurement, the stimuli were two "roads" composed of match sticks of unequal length placed end to end. Two sets of sticks were initially lined up parallel to each other, and without overlap, to form two equally long lines. The *S* was asked whether the "roads" were the same length or not. Even the youngest *S*s were able to respond "Yes" to this question.

One line of sticks was then altered so that the "road" was no longer straight, but formed a "zig-zag." The *S* observed the alteration. He was then asked again whether the "roads" were the same length or not. This question was sometimes rephrased, the child being asked, if an ant were to walk along each road, whether one ant would walk farther than the other.

It was found that the youngest children (aged 4–6 yr.) did not consider the lines the same length after the rearrangement had taken place; they considered either the straight line to be longer because it extended beyond the other one, or else they thought the "zig-zag" line to be longer because of the turns. There was an intermediate stage where some rearrangements were considered to alter the length and some were not. Around ages 7 and 8 yr. (Stage III), the line was known to be the same length irrespective of the rearrangement of its parts.

This experiment is one of the rare ones in which Piaget reports some results numerically. Fifty-nine children between the ages of 4 and 9 were examined. Responses classified in Stage III were shown by only 10% of those aged 6–0 to 7–0, by 50% of those between 7–0 and 7–6, and by 75% of those between 7–6 and 8–6.

No manipulation of measuring instruments is required of the child in this experiment, but Piaget's questions are very complex for young children. Indeed, the use of words such as *long* or *length* (which appear in the questions) to refer to lines which are not straight presupposes comprehension of the additive operation involved in measurement, since the length of a line which is not straight is defined as the sum of the lengths of its straight parts. Comprehension of Piaget's questions by the child would therefore of itself indicate understanding of the additive character of lengths. Piaget's results can thus only be regarded as equivocal, since it is unclear precisely what he is studying. The development of this additive operation could perhaps be examined more adequately if the experiment were redesigned as a length-discrimination task in which selection of the longer of two sticks (at least one of which is bent to form a zig-zag) led to, say, a candy reward. At whatever age the additive operation develops (if there is such an age) one would expect an increase in skill in visualizing a line as composed of displaceable parts, and therefore a sudden improvement in the accuracy of discrimination at this age.

Vocabulary development may well be a factor in many of Piaget's experiments. Its confounding effects cannot be eliminated unless nonverbal methods of presenting the tasks to the *S*s are used. The notion of employing nonverbal techniques has methodological importance. Historically, one of the important contributions of comparative psychology has been the consistency with which processes of reasoning or concept formation, which at one time were considered by some to be peculiar to the human species, have been elicited in lower animals, once the appropriate nonverbal tasks have been designed. Field's (1932) study of concept formation in the rat provides an example of this. The success of Harlow (1951), and Evarts and Nissen (1953) in teaching subhuman primates to solve conditional matching problems employing the Weigl principle is a more recent instance, which suggests that the kind of thinking called "abstract" by Goldstein and Scheerer (1941) can occur in primates other than man.[8]

No theory which postulates levels of conceptual development can be regarded as definitively established when the supporting data are obtained through extensive verbal communication with *S*s who differ in their ability to verbalize. This principle applies not only when the human species is compared with phylogenetically lower animals, but also, within the human species, when aphasics—and perhaps also schizophrenics—are compared with "normals," and in ontogenetic studies of young children.

The main purpose of using nonverbal techniques is therefore not to discredit the notion of "levels" or "stages" of development, but to provide more convincing evidence for them.[9]

Conservatively interpreted, the evidence suggests the following conclusions:

1. In the absence of measurement, the accuracy of length judgment in young children varies with the nature and the spatial arrangement of the stimuli judged. Where the lengths are the paths traced by moving objects, failure to notice or to recall the relative starting positions is an important source of error.

2. Selection and manipulation of instruments for measuring length becomes increasingly adequate between the ages of 4 and 8 yr. in the average child who has had no special tuition.

3. There is reason to question Piaget's belief that performance of the inferences and logical operations basic to measurement emerges at the approximate age of 7 yr. in the average child. Piaget's findings may reflect his methods of investigation.

To evaluate this criticism of Piaget, a length-discrimination task was designed in which the difference in length between the stimuli could not readily be perceived but had to be inferred, using the rule $A > B.B > C. \supset .A > C$. This rule has the status of an axiom in the logic of measurement. The *S* did not manipulate any measuring instruments during the task so that skill in the use of measuring rods was not a controlling factor. Every time the child made a correct response he obtained a small piece of candy, so that he was not motivated only by the inherent interest of the task.

Tracing the development of this inferential response in children makes it possible to subject to empirical test Piaget's view that the inferences fundamental to length-measurement develop in children at around age 7 yr.

Method

Subjects

The *S*s were children from two day nurseries in New York City, partially supported by the Division of Day Care of the New York City Department of Welfare. There were 41 *S*s, 18 boys and 23 girls, who ranged in age from 3–6 to 7–0. The distribution by age is shown in Table 1. The children in these nurseries are almost all children of working mothers who live in low-income housing projects in New York's lower East Side. Although the socioeconomic status and education of the parents are below average for New York City, in the opinion of the nursery authorities and of the writer, it is doubtful that the children are below the country-wide average in general intelligence. IQs are available for very few of the children, but those that are range from 89 to 118 and thus are in harmony with this general impression.

Of the 41 *S*s, 37 also received the order-discrimination task described in the next section. To 16 of these, the order-discrimination task was administered before the task now described.

---

[8] Moreover, deficits in "abstraction" following brain injury may not be the peculiar privilege of man! Cf. Riopelle, Alper, Strong, and Ades (1953).

[9] The writer is indebted to H. L. Teuber for suggesting the use of nonverbal techniques. In numerous conversations he has persistently reiterated that, if a theory postulates an alteration of thought (rather than language) processes either during development or as a result of brain injury or disease, then the possibility of demonstrating this alteration on nonverbal tests is clearly implied.

Apparatus

The apparatus consisted of 15 upright pieces of wood painted a turquoise color, screwed to a flat base painted black. Each of these could be placed to rest on a low orange pedestal, constructed so that a piece of candy could be concealed in the pedestal beneath the base of the upright. The apparatus also contained two orange measuring sticks. The 15 uprights are shown in Fig. 2. Two of these 15 are shown in Fig. 3 being measured in turn with a measuring stick. The longer of the two measuring sticks, which measured $11\frac{3}{8}$ in. and $11\frac{3}{4}$ in., is shown in Fig. 3. Each of the measuring sticks was designed with a spigot on top which passed over the shorter of the two uprights, thus clearly revealing the relative lengths of the uprights and measuring stick. This spigot can be seen best in the left photograph of Fig. 3.

The purpose of constructing the uprights with arms at their upper and lower extremities was to increase the difficulty of making accurate discriminations of their relative lengths by direct visual comparison. It was expected that the arms would induce a slight degree of illusion (Müller-Lyer effect) in the apparent length of the uprights.

Procedure

The task consisted of a series of discrimination problems with the stimuli (the uprights) presented in pairs. There were three phases. In the first phase the uprights used were always clearly different in height. Half the children learned to obtain a reward by selecting the longer upright, the remainder by selecting the shorter upright. In the second phase *S* was required to find the rewarded upright when the two uprights in each trial were

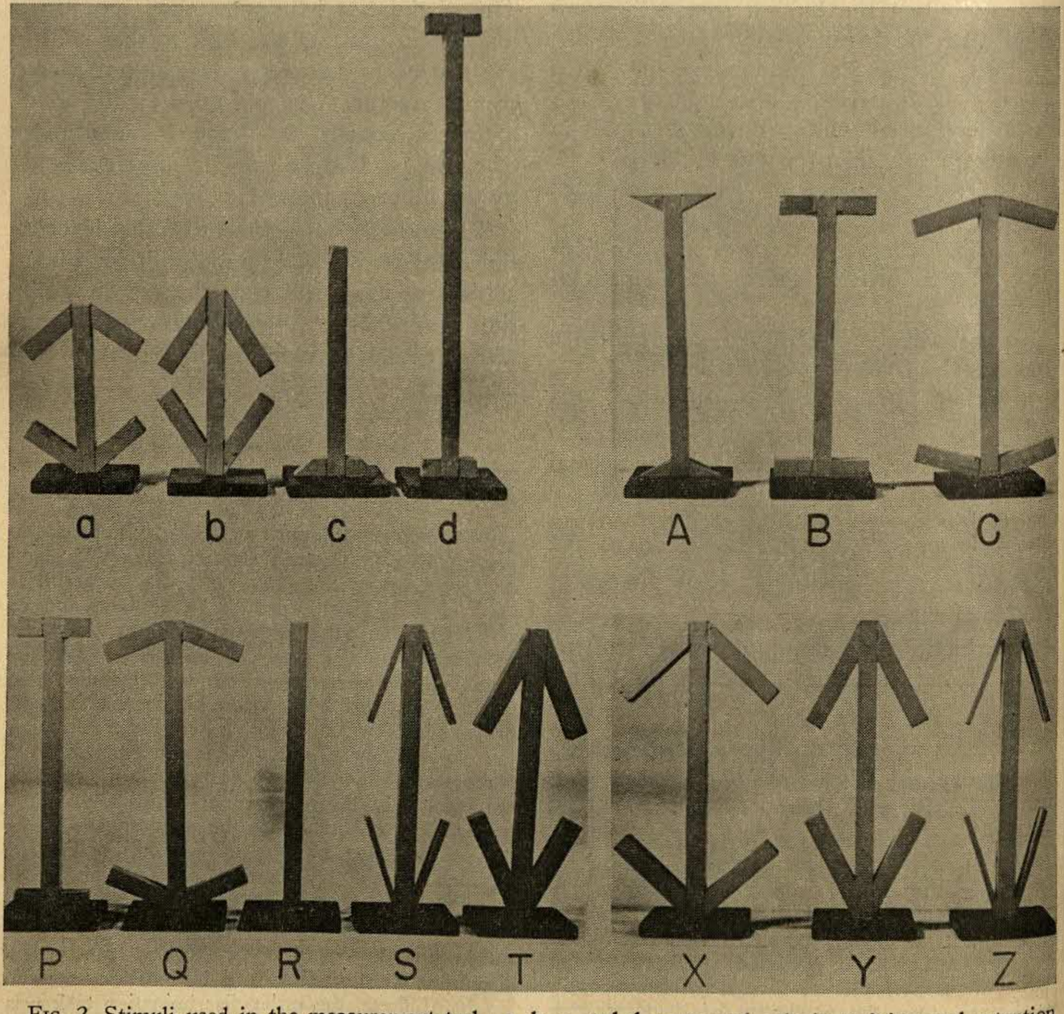


FIG. 2. Stimuli used in the measurement task. a, b, c, and d were used only in training and retention trials. A, B, and C were $11\frac{3}{16}$ in. tall, P, Q, R, S, and T $11\frac{9}{16}$ in., and X, Y, and Z $11\frac{15}{16}$ in.

objectively, but not discriminably, different in length, and where each was compared in turn, by *E*, with a measuring stick of intermediate size. The *S* had to *infer* which was the longer (or shorter) upright, by the rule $A > B . B > C . \supset . A > C$. In the third phase the question, "Which is the taller [or shorter]?" was posed verbally as *E* measured the uprights with the measuring stick. Thus, in Phase 1 each child was taught the relation "longer than" (or "shorter than"). Phases 2 and 3 tested whether this relation was transitive or intransitive for the child: in Phase 2 a nonverbal method was used, and in Phase 3 the nonverbal method was supplemented by verbal instructions. The precise details of procedure in the three phases were as follows.

Phase 1: Training. The *S* and *E* sat opposite each other, with the long edges of a low rectangular table between them. The *S* was first shown how to lift one of the uprights and find candy under it. The uprights were then presented in pairs. Half the children in each age group always found candy under the taller upright, the remainder under the shorter upright. The uprights in each pair were clearly different in height: during Phase 1 no two of the uprights designated in Fig. 2 by capital letters were ever paired together. The uprights designated d and P in Fig. 2 constituted the first pair of training stimuli. This pair was presented in successive trials until *S* had achieved eight successive correct comparisons. The pair d-P was then removed and a-X substituted for it. When *S* had performed six successive correct comparisons of a and X, these were removed and d-P reinstated. If *S* did not remember whether the candy was under d or P, he was retrained. Pairs d-P and a-X were then presented on successive trials until both discriminations were retained simultaneously. The pairs a-P, P-d, d-X, and X-a were then presented in successive trials until all four discriminations had been learned and retained simultaneously. If *S* proved unusually slow in learning to transpose the concept taught, these pairs were presented individually in successive trials until learned. Then c-R was introduced and *S* was trained on this if necessary. Then the pairs b-c, c-R, R-d, d-Y, Y-c, c-a were presented successively until all the discriminations involved were retained simultaneously. When this point was reached, uprights designated by lower-case letters in Fig. 2 were paired at random in successive trials with uprights designated by capital letters until eight successive correct comparisons of different pairs of uprights were made. The *S* was then considered to have learned whichever of the concepts "longer than" or "shorter than" he had been taught and Phase 2 was initiated. When the training phase required more than one session with an *S*, pairs of stimuli were presented in subsequent sessions in the same sequence as they had been presented in the first sessions: that is, the child was always trained from the beginning again, although the success criteria for the discriminations learned in prior sessions were somewhat relaxed.[10]

During Phase 1, the position of the correct upright was varied, but it was not always varied randomly. It was found that random variation of the correct position was an ineffective training technique since it generally provided accidental partial reinforcement of an *S*'s wrong hypotheses or habits. Many *S*s tended to reach for the upright nearest their preferred hand. It was found that the only effective way of countering this was to put the correct upright consistently on *S*'s nonpreferred side until the position habit was abandoned. In general, during Phase 1 a random position sequence was employed from which *E* departed periodically in order to avoid reinforcing incorrect hypotheses that *S* appeared to be developing.

At the beginning of the training in Phase 1 the uprights were placed on the table a few inches apart in front of *S*. They were gradually moved away from each other until one was on *S*'s extreme left and the other on his extreme right:

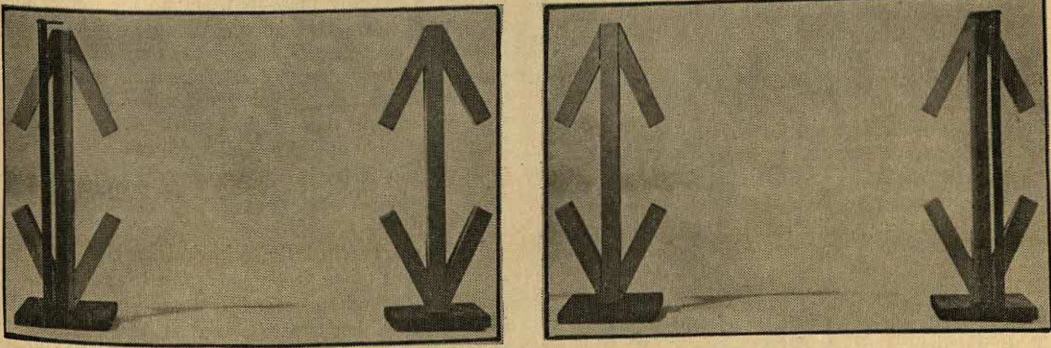


FIG. 3. Example of a measurement trial: the measuring stick is juxtaposed with each upright in turn.

---

[10] The success criteria were selected following Hilgard (1951, Table 5, p. 535) which gives the number of successive correct responses required for statistical significance as a function of the total number of trials.

these were the positions the uprights occupied in Phase 2, in order to render direct visual comparison of their heights more difficult.

Phase 2: Measurement and retention. This phase consisted of a minimum of 40 trials. Eighteen of them were "measurement" trials. Ten were "trials without yardstick" and there were at least 12 retention trials.

The uprights were always presented in pairs as in Phase 1, and the candy was always under the same upright (the "longer" or the "shorter") as it had been in Phase 1. In the retention trials the same combinations of stimuli were used as had been used in the training trials of Phase 1. However, in the measurement trials and trials without yardstick only uprights designated in Fig. 2 by capital letters were used. The objective differences in height between the stimuli were therefore always quite small. It will be noted that among the 11 uprights designated by capital letters in Fig. 2, there are 39[11] possible pairings in which there is a small true difference in height. In a few pretesting sessions the 28 pairs which seemed most difficult to discriminate by direct visual comparison were selected and 18 of these were randomly assigned to the measurement trials. The remaining 10 pairs were used in the trials without yardstick.

During measurement trials *E* juxtaposed one of the measuring sticks with each upright in turn, the left-hand one first, as shown in Fig. 3. The measuring stick was always shorter than one of the uprights and longer than the other. It was placed vertically on the base of each upright and rotated so that the spigot passed over the top of the shorter upright or rested against the side of the longer upright (see Fig. 3). If *S* did not seem to be watching during this procedure, the stick was knocked against the top of the upright to attract his attention and he was told, "Look at the tops of them." The *S* was not permitted to touch either of the stimuli until the measurement had been performed. The measuring operation was often prefaced by the words, "Wait, we are going to try and help you find the candy." If the child failed to lift one of the uprights immediately after the measuring operation had been performed, this operation was repeated and then he was instructed, "Find the candy." The criterion for successful "measurement" was 13 correct trials out of 18.

The trials without yardstick were the same as the measurement trials except that the entire measuring procedure was omitted. The *S* was instructed, "See if you can find the candy without help this time." The purpose of the trials without yardstick was to test the accuracy of *S*'s discrimination of the lengths of the uprights without measurement, in order to show that in the measurement trials *S* had "inferred" rather than directly perceived the longer (or shorter) upright.

In the retention trials pairs of uprights clearly different in height, as in Phase 1, were used. The retention trials were included in order to ensure that *S*s who were unable to find the correct upright in the measurement trials still retained the concept "longer" (or "shorter"), i.e., to ensure that failure on measurement trials did not occur simply because *S* had forgotten the problem "instructions," taught in Phase 1. When failure occurred on any of the retention trials the original length discrimination was retaught as in Phase 1 to the same criterion of eight successive correct comparisons of eight different pairs of stimuli, before the measurement trials and trials without yardstick were continued. No *S* was retrained more than three times.

The measurement trials were given in pairs, and between every pair of measurement trials a trial without yardstick was inserted. The retention trials were distributed among the 28 measurement trials and trials without yardstick: between every fourth and fifth of these trials, two retention trials were inserted.

Throughout Phase 2 the position of the correct upright was varied randomly.

Phase 3. The procedure in this phase depended on whether *S* had reached the criterion of success in the measurement trials of Phase 2. If *S* had failed to reach this criterion, the following procedure was adopted. Two uprights similar in appearance (e.g., S and Z or T and Y in Fig. 2) were first placed next to each other and inclined together until the tops touched. The *S* was told to put his finger on the top of the one that was "taller or bigger" (or, "shorter or smaller"). The *S* then watched a piece of candy being put under the taller upright (or the shorter, whichever concept had originally been learned). The positions of the uprights on the table were then rapidly interchanged several times so that *S* lost track of the one with the candy. The uprights were then measured as in the Phase 2 measurement trials. As each upright was being measured, *S* was instructed to point to the "one that is longer or taller" (or, the "one that is smaller or shorter"). Thus, an *S*, for whom the longer upright had previously been rewarded, now pointed to the measuring rod as the shorter upright was being measured, and to the longer upright as this was measured. When both had been measured, he was told, "Find the candy under the one that is longer or taller." The main purpose of the supplementary verbal instructions in Phase 3 was to ensure that *S* had noticed the relative lengths of the uprights and measuring rod. Not more than six trials were given. To be credited with success, *S* had to respond correctly in five successive trials.

If *S* had reached the criterion on the Phase 2 measurement trials, it was desirable to make doubly sure that he had succeeded by a genuine process of inference. A book was placed under one of the pedestals so that the bases of the pedestals were no longer at the same level. Two uprights similar in appearance (e.g., S and Z or T and Y) were again selected and placed one on each pedestal. The child was asked which was the taller (or shorter, if he had originally learned this discrimination). Following his answer the uprights were measured in the usual manner, and he was then asked whether his original judgment had been correct or not. The uprights were reshuffled several times and this procedure repeated until *S* had demonstrated without error in 5 trials that he was able both to maintain and to change his original judgment, depending on the facts as indicated by measurement.

---

[11] Fifteen pairings of A, B, C with P, Q, R, S, T + 9 pairings of A, B, C with X, Y, Z + 15 pairings of P, Q, R, S, T with X, Y, Z.

Results and Discussion

Level of Performance Attained

One *S* failed to reach the learning criterion in Phase 1 and was therefore excluded from the subsequent phases.[12] The remaining 40 *S*s all learned the original length discrimination in Phase 1, and to all of these *S*s Phases 2 and 3 were therefore administered.

Fifteen *S*s failed to reach the criterion of 13 correct measurement trials in Phase 2, and also failed in Phase 3. These *S*s therefore failed to find the longer or shorter upright by inference in both phases. An additional 5 *S*s failed to reach the criterion in the measurement trials of Phase 2, but made an inferential response in Phase 3, when the nonverbal method was supplemented by verbal instructions. The remaining 20 *S*s were successful under both conditions; there were no *S*s who reached the criterion in Phase 2 but failed to do so in Phase 3.

The percentage of *S*s within each 8-mo. age interval between 3–6 and 6–10 who successfully found the correct upright following the measuring procedure is shown in Fig. 4. This graph shows separately the increase with age of the percentage of *S*s reaching the criterion in the measurement trials of Phase 2, and those who made an inferential response in either Phase 2 or Phase 3. It can be seen that a regular development appears to occur, and that in this sample the threshold age at which 50% of children made the inference studied is somewhere between 4–2 and 5–5. In Table 1 the development with age is shown in terms of the mean number of correct responses in the 18 measurement trials.

The improvement in performance with age in the trials without yardstick is also shown in Fig. 4 and Table 1. A criterion of eight correct choices in the 10 trials without

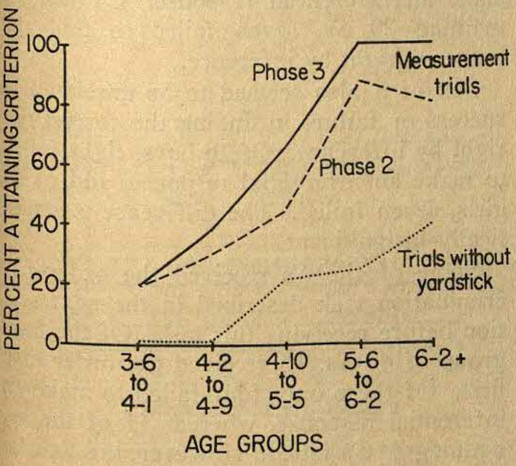


FIG. 4. Percentage of *S*s finding the correct upright by inference (measurement trials) and by direct comparison (trials without yardstick).

TABLE 1

NUMBER OF CORRECT RESPONSES AS A FUNCTION OF AGE IN THE MEASUREMENT TRIALS AND TRIALS WITHOUT YARDSTICK OF PHASE 2

Age Group	*N*	Mean No. Correct Responses	
		18 Measurement Trials	10 Trials Without Yardstick
3–6 to 4–1	5	9.5	5.0
4–2 to 4–9	13	10.2*	4.6
4–10 to 5–5	9[a]	12.4****	6.0
5–6 to 6–1	8	14.8****	5.7
6–2 to 7–0	5	15.6****	7.2**

Note.—The statistical tests were based on the null hypothesis that *S*s' scores are distributed as predicted by the binomial expansion, with N = no. of trials and $p = q = .5$. Thus, where n = no. of *S*s in an age group, z = (Mean Score of age group $- Np) / \sqrt{Npq/n}$.

[a] One *S* excluded who failed to learn the initial length discrimination in Phase 1.

* Significant at the .05 level.

** Significant at the .01 level.

**** Significant at the .0001 level.

---

[12] This child was aged 5–0. In three training sessions she failed to learn even the first discrimination (between P and d). The writer has no explanation for this failure.

yardstick was used in Fig. 4.[13] It can be seen that in all except the oldest age group the average number of correct responses in the trials without yardstick is close to the chance level. Even in the oldest age group accuracy is considerably less than in the measurement trials.

Since the use of the measuring stick is the only factor that distinguishes the measurement trials from the trials without yardstick, the greater accuracy in the measurement trials is attributable to the *S*s' use of the cues provided by the measuring procedure. Further evidence that the children successful in the measurement trials found the longer or shorter upright by inference is provided by the fact that all these children were able to find the correct upright in Phase 3, when the bases of the uprights were at different levels.

The evidence that the children successful in the measurement trials were using an inferential procedure is thus convincing. The results suggest also that the lack of success of the children who did not reach the criterion in either the measurement trials of Phase 2 or in Phase 3 cannot be explained simply as a failure to retain the original concept "longer" or "shorter," taught in Phase 1. Evidence for this is found in the retention trials of Phase 2. Of the 15 children who failed in both the measurement trials of Phase 2 and in Phase 3, six responded correctly in all 12 retention trials, five failed in one retention trial and had to be retrained once, and three had to be retrained twice. One *S* was discontinued on Phase 2 because of consistent failure to retain the discrimination. Of the 20 children who reached the criterion on the measurement trials, three had to be retrained once and one twice. Thus, *S*s who failed to make an inferential response also tended to forget the discrimination taught in Phase 1 somewhat more often than *S*s who made an inferential response. This is not surprising since the former *S*s were in general younger and more distractible. But it is extremely unlikely that failure to reach the criterion in the measurement trials is due to a simple failure to retain the original discrimination. It may be recalled that the retention trials were spaced among the measurement trials and trials without yardstick in such a way that any failure to retain the discrimination would be very quickly discovered. If this happened, *S*s were then retrained before any further measurement trials or trials without yardstick were administered. Thus one or two failures of retention would not vitiate the results in the measurement trials of Phase 2. Moreover, an *S* is not considered to have failed to have made an inferential response unless he also failed in Phase 3 with the supplementary verbal instructions.

Three variables probably unrelated to successful "measurement." Whether or not an inferential procedure was used seemed unrelated to the location of the candy under the longer or shorter upright. For 20 *S*s the reward was consistently under the shorter upright, and of these, eight failed to make an inferential response. Of the remaining 20 *S*s, seven failed to find the longer upright by inference.

Sex of *S* also seemed to be unrelated to success or failure in finding the correct upright by inference. Of 18 boys, eight failed to make an inferential response, and of 22 girls, seven failed. The difference is statistically insignificant.

Some of the *S*s received the order-discrimination task described in the next section before receiving this task. Of the total group of 40 *S*s, 16 received the order task first. Of these only four failed to make an inferential response, whereas 11 of the remaining 24 *S*s failed. However, the *S*s who received the order-discrimination task first were generally older. This was because a few young *S*s inopportunely went on vacation after receiving the order task. Unlimited time was not available and it was therefore desirable to begin the measurement task early, since, owing to the lengthy Phase 1 training, it required several sessions to administer to the younger children. Of the 21 children under 5-0, 14 took the measurement task first and 10 failed to make an inferential response. Of the seven who took the order task first, four failed to make an inferential response. Chi square, even without Yates correction, is very small; even for the whole group of 40 *S*s chi square is not significant. Thus there is no evidence that prior experience with the order task affected the *S*s' success in making an inferential response in the measurement task.

---

[13] This criterion errs on the lenient side: $p =$ .054 of obtaining by chance eight or more correct responses in 10 trials. This criterion was selected because it is comparable to the criterion used in the measurement trials ($p = .048$ of finding 13 or more correct in 18 trials).

Temporal course of performance in Phase 2. In Fig. 5 the temporal course of performance in the 18 measurement trials is shown separately for the group of 20 *S*s who reached criterion and for the remainder who did not. It can be observed that there was no tendency to learn on the part of *S*s who failed; their performance never deviated far from the chance level. The 20 *S*s who reached criterion performed significantly better than chance even on the first three trials, and the improvement that followed took place very quickly. Moreover, even on the first measurement trial, 16 of these *S*s found the correct upright.[14] The evidence therefore indicates that *S*s who reached criterion on the measurement trials tended to use the inferential procedure from the beginning of Phase 2.

In Fig. 6 the course of performance on the 10 trials without yardstick is shown. It is noteworthy that the *S*s who reached the criterion on the measurement trials performed consistently better than the remainder. That is, the *S*s who made an inferential response also tended to discriminate the relative lengths of the uprights better when no inferential procedure was involved. In the last four trials without yardstick their performance even rose to a level significantly better than chance.

This correlation raises some very interesting questions. In particular, it might lead one to suggest that the dichotomy originally made between "inferred" and "perceived" length differences may not be as absolute psychologically as it is logically. This question is complicated and a discussion of it is deferred until it can be treated within the context of the totality of the results.[15] At this point one may note only that the *S*s who made an inferential response were older. Also, in the specific stimulus conditions used in this experiment, discrimination of the lengths of the uprights without the aid of the measuring stick is quite complex, because of the slight degree of illusion

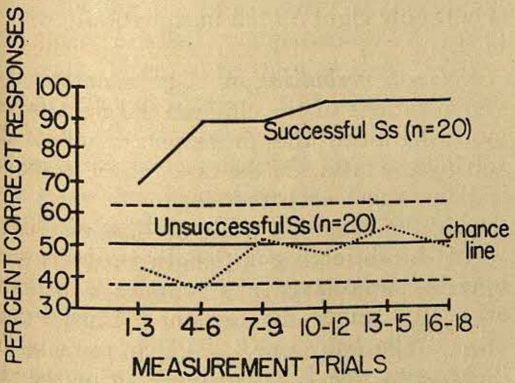


FIG. 5. Temporal course of performance on the measurement trials. The "successful" *S*s are those who reached the criterion. The parallel broken lines enclose the region between the minimum score significantly greater than chance and the maximum score significantly less than chance ($p = .05$).

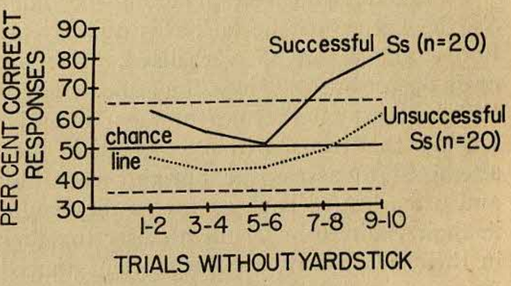


FIG. 6. Temporal course of performance on the trials without yardstick. The "successful" *S*s are those who reached the criterion in the measurement trials. The parallel broken lines enclose the region between the minimum score significantly greater than chance and the maximum score significantly less than chance ($p = .05$).

---

[14] $p < .05$ that this is a chance occurrence.

[15] See section entitled Changes in "Set" Associated With the Emergence of Measurement.

presumably induced by the arms at the upper and lower extremities of the uprights.[16]

Qualitative results. For the purposes of description, the *S*s' behavior may be grossly categorized into three types. However, there were many *S*s who did not fall clearly into any type. Also, at the beginning of the experiment it was not known how they would behave. On looking back at the protocols, it was found that for some of the *S*s run first there was no adequate record of the presence or absence of many of the behaviors later observed carefully.

The *S*s who failed to make an inferential response appeared to have no comprehension of the measuring operation in Phase 2. They did not spontaneously watch the measuring procedure except on the first few trials. Their attention had to be constantly attracted to the measuring sticks by various means, with the result that administering Phase 2 so that the measurement was at least observed, if not understood, proved to be quite exhausting to *E* and probably to *S* also, and often required several sessions. On the trials without yardstick these *S*s did not seem to compare the uprights visually. They gave no more than a quick glance at each and then reached for one of them. It was as if the possibility of an imperceptible difference in length was beyond their comprehension.

To others, who were generally the older *S*s, the nature of the task was quite clear. For example, one *S* verbalized, "I know one's bigger but they look the same." These *S*s would watch the measurement intently. During the trials without yardstick they attempted to assess the uprights carefully and often leaned back as far as they could in their chair to bring them closer together in their field of vision; they often glanced from one upright to the other several times. In general, these *S*s found the measurement trials rather easy and one even urged *E* not to use the measuring sticks because that "gave it away."

There was a very interesting third group, into which fell many of the younger *S*s who made an inferential response. Like the first group, they did not spontaneously watch the measuring procedure in Phase 2, but when their attention was drawn to it, they proved able to find the candy with the aid of the measuring stick. Some of these *S*s might find the candy two or three trials in a row, while paying attention, and still fail to watch spontaneously the measuring procedure in subsequent trials. Their behavior recalls the observations of Goldstein and Scheerer (1941) in which they argue that abstract thinking is an "attitude" rather than an "ability"; these *S*s were able to make an inferential response, but did not spontaneously attempt to do so. Some of the *S*s in this group reached the criterion in the measurement trials of Phase 2, whereas others only demonstrated their inferential ability in Phase 3, where the verbal instructions permitted *E* to take a more active role in fixing the *S*s' attention. Like the *S*s of the first group, these *S*s showed no comparison behavior in the trials without yardstick of Phase 2. This group was small: only eight *S*s fell unequivocally within it.

Role of verbalization. The spontaneous verbalizations of the children did not throw much light on the processes involved in solving the task. Of the task-related verbalizations, most occurred during Phase 1 or during the retention trials of Phase 2 (i.e., when the difference in length between the uprights was large). Examples of these are: "It's under the big one," "The little one," "The baby one," "Which one's up," "The teeny one." Interestingly, it appeared that when an *S* says, "It's under the big one," this should not necessarily be construed by *E* to mean "It's always under the big one." An *S*'s verbalization often seemed to refer only to a particular trial, and he might select the incorrect stimulus on the very next trial.

There was no evidence that verbal cues were of much help in solving the measurement trials. Spontaneous verbalization during Phase 1 of the location of the candy under the longer or shorter upright was not related to the performance of an inferential response. No child spontaneously verbalized the measuring procedure. At the end of the task several *S*s were asked how the use of the measuring stick helped them find the candy. All except two said they did not know. Of these two, one answer was, "Because it [the stick] makes it [the upright] smaller." The other answer was, "If the thing [measuring stick] is bigger than there [upright] and the other one the stick's smaller, it's there." Both these *S*s were girls aged 5-6. The latter child was the one, cited earlier, who verbalized, "I know one's bigger but they look the same." The performance of this *S* was interesting. During the measurement trials she pointed occasionally to one or the other of the two uprights, as if she were trying to verbalize the solution of the problem to herself subvocally. These efforts seemed to be more confusing than helpful to her: in the 18 measurement trials, she failed the first trial, solved the next six, then made four successive errors, followed by seven successive correct responses. She did not become similarly confused in the trials without yardstick, finding the objectively taller upright after careful visual comparison in 9 of the 10 trials. Even when the question "Which one is taller or bigger?" was posed verbally by *E* in Phase 3, she almost became confused again but managed to correct herself. She gave the impression that she was trying to form the series long-upright—measuring stick—short-upright, but, having formed the series, she forgot which end was which!

---

[16] It will be recalled that the uprights were deliberately constructed in this manner in order to increase the difficulty of discriminating their relative lengths without the aid of the measuring sticks. If the *S*s had shown a moderate or high degree of success in the trials without yardstick, it would have been difficult to prove statistically that *S*s who were successful in the measurement trials had used an inferential procedure.

In general, although the location of the candy under the longer or shorter object was often verbalized, there was no evidence that any verbal cues were used which related specifically to the inferential procedure: no audible words were used in the measurement trials which were not also used in the initial learning and in the trials without yardstick. In the one *S* who appeared to be seeking to use verbal cues during the measuring procedure, these cues seem to have been more confusing than helpful.

Significance of Results for Piaget's Theory

The most important single result of this study is the finding that the inferential response $A > B . B > C . \supset . A > C$ can be elicited in children substantially earlier than Piaget leads one to believe. If Piaget's finding that children do not manipulate measuring instruments effectively until around age 7 or 8 yr. is valid, then children can perform this inferential response long before they are able to measure adequately. While the availability of this response may be a necessary condition for adequate measurement, it is therefore not a sufficient condition. One plausible reason for the apparent time lag between the development of the inference studied in this experiment and the development of adequate measuring behavior in Piaget's research is that manipulation of measuring instruments requires considerable learning of specific measuring techniques. Adequate measuring behavior may also demand interest in measurement and in achieving the additional accuracy that it makes possible. It is likely, therefore, that many of Piaget's studies, which he believes investigate the development of children's performance of the logical operations implicit in measurement, should be reinterpreted as studies of the development of skill and interest in the techniques of measurement. Inferential responses of the kind investigated here may be available to many of his *S*s, particularly those of intermediate age.

The experimental technique used also differed from Piaget's in the manner in which the stimuli were presented. Piaget used instructions which contained the words "same length," and "measure," or synonyms of these. Until Phase 3 at least, the task was presented nonverbally, except for the instruction, "Find the candy"; moreover, the words "longer" or "shorter," introduced in Phase 3, were preceded by the considerable training in length-discrimination involved in the previous phases.

Probably this difference in technique is a factor contributing to the difference between the present results and Piaget's. It is particularly likely that the nonverbal method facilitated the success of the youngest children making an inferential response.

The fact that the findings of this experiment indicate that Piaget is mistaken about the specific age at which the inferences basic to measurement develop raises the question: Is it possible—and if possible, is it desirable—to attempt to decide at what age these inferences develop in children? In this sample of children, presumably of average intelligence, and under the particular stimulus and response conditions employed here, this age was found to be somewhere between 4-2 and 5-5. However, it would be incautious to assert that this is the earliest age at which these inferences can be elicited in average children. Although a considerable effort was made to use experimental operations simpler than Piaget's, it is quite possible that further study would lead to the discovery of procedures in which the inferential response studied could be elicited in still younger children, thus demonstrating that some relevant factor was overlooked in the present experimental design.

The crucial theoretical issue concerns the usefulness of postulating the existence of an "emergent" formal reasoning factor in the development of children's length judgment.

It could be argued that if the apparent "age of emergence" is a function of the particular manipulanda and discriminanda involved and of their mode of presentation to *S*, then the concept of an emergent level is unsatisfactory in that it lacks parsimony and could even obstruct psychological analysis. One might be forced to postulate as many "levels" as there are different experimental operations for eliciting the same alleged psychological process.

One weakness of this argument against the "levels" concept is that it could clearly be transposed and used to attack, not only Piaget's, but any concept of emergent levels. Moreover, one can argue that there must be some point between birth and maturity at which the kind of inferential process studied in this section develops in the average child. The reason why Piaget's results differ from those reported here is not simply that there is a difference in experimental procedures, but is, the writer believes, because in designing his experiments Piaget fails to eliminate important variables which are not involved in the definition of the processes he sets out to investigate. That is to say, in his studies of measurement, Piaget claims to be studying the development of the formal reasoning inherent in measurement; he is probably actually studying primarily the development of skill or interest in the use of measuring rods, or perhaps the development of children's understanding of certain words. The results obtained here therefore lend support to the notion of an "emergent" reasoning process in the development of children's length judgments, although the establishment of "emergent levels" is a more complex experimental problem than Piaget apparently seems to believe.

It is clear that if one seeks to state an age at which a particular type of response develops, the only age which is not completely arbitrary is the earliest age at which this type of response can be elicited using the simplest experimental procedures. To establish these "simplest" procedures requires the discovery and systematic exploration of all relevant factors. The experiment reported can be interpreted as a step in this direction. Such a program of research would presumably result either in a clear operational definition of the alleged "emergent" process, or in convincing evidence that the notion of an "emergent" reasoning process in children's conception of length is unsatisfactory because no clear distinction between the "higher" and "lower" processes can be found. There is nothing in this study to suggest that the formal reasoning inherent in measurement cannot be defined experimentally.

Changes in "Set" Associated with the Emergence of Measurement

In presenting the qualitative results, differences in behavior were described, which seemed broadly related to age. The oldest *S*s showed the most comparison behavior, especially in the trials without yardstick: they tended to lean back in their chairs to facilitate visual comparison of the uprights; they paused before responding and looked intently at each upright, often giving each several glances. Such behavior indicates that they readily transferred the set to discriminate length, learned in Phase 1 with the large and obvious length differences, to Phase 2 where the length differences were small. In contrast to this, comparison behavior was conspicuous by its absence in many of the youngest *S*s. This fact suggests that perhaps differences in length too small to be readily perceptible did not have much significance for the younger children. It was as if they were not "trying" to find the longer upright when the differences were small, not because they were unmotivated, but because they failed to recognize the possibility of a difference in length too small to be noticed. They seemed to lack the attitude of the older *S*s expressed in the words, "I know one's bigger, but they look the same."[17]

Thus, there is evidence that two kinds of behavior develop. First, there is the development of the inferential response, which is revealed quantitatively in the increasing success with age in the measurement trials of Phase 2. Second, there is a change with age in the "attitude" or "set" adopted toward small differences in length, which seems to consist in a greater tendency with increasing age to transfer the set to discriminate length from large obvious differences to small barely perceptible ones. This change in "attitude" is revealed quantitatively in the better-than-chance success of the oldest *S*s in the trials without yardstick.

The question must be raised whether there is any connection between these two phenomena. Is the development of the inferential response related to the phenomenon referred to as a change in attitude? On this issue there are three possible hypotheses which are worth exploring, although the present data do not clearly indicate which hypothesis is valid.

The first hypothesis is that the change in "attitude" or "set" and the inferential response develop quite independently and are unrelated in the individual child. If this is so then there should have been no tendency for the *S*s who made an inferential response to have done better in the trials without yardstick than those who failed to make an inferential response. In the total sample this was certainly not true. The relationship can be clearly seen in Fig. 6. However, when age was taken into account the correlation was equivocal. Matching the seven youngest *S*s who made an inferential response with *S*s of the same age who failed to do so, it was found that those who made an inferential response obtained a mean score on the trials without yardstick significantly higher than the others ($t = 2.47$, $df = 6$, $p < .05$); however, both means (5.7 and 4.4) can themselves be readily explained on the hypothesis that the *S*s were responding at random. Consequently the evidence against the first hypothesis is ambiguous. The remaining hypotheses both assume a relationship.

The second hypothesis is that the change in attitude is the determining variable and that the development of an inferential response is dependent on it. That is, the adoption by *S* of the more "mature" set is a precondition for the occurrence of an inferential response: unless *S* realizes the uprights differ in length and consequently seeks the longer, he will not infer the longer upright in the measurement trials. Plausible though this hypothesis may be, the evidence is against it. The best evidence is qualitative. Although all *S*s who showed comparison behavior also made an inferential response, there were at least eight *S*s who made an inferential response but who showed no comparison behavior in the trials without yardstick: there was no change in posture to bring the uprights closer together in the field of view, the response was not preceded by several ocular shifts from one stimulus to the other, and there was no verbalization which suggested that these *S*s realized the stimuli differed in length. Thus, an inferential response could occur in the context of behavior associated with the more "primitive" set. The quantitative evidence of Fig. 4 points in the same direction: inferential responses occurred in much younger *S*s than did any success in the trials without yardstick. Unfortunately, however, the quantitative evidence is inconclusive because of the illusion presumably induced by the arms placed at the upper and lower extremities of the uprights: the possibility cannot be dismissed that the younger *S*s failed to discriminate to a degree better than chance because they were deceived by the illusion, and not because their "set" was inappropriate.

---

[17] If this interpretation is correct, it raises the more general problem of how and when children learn to make the conceptual distinction between a "real" and a "phenomenal" world. This problem has been largely ignored by experimental genetic psychology outside Geneva. No general discussion of it is attempted here.

The evidence, although primarily qualitative, is more consistent with the third hypothesis, which is that the change in attitude toward barely perceptible length differences is dependent upon the prior development of inferential processes of the type studied. Should this hypothesis prove valid, it could be explained on the theory that the development of the inferences basic to measurement leads to the establishment of a concept of length as a dimension in which continuous small increments are possible. A frame of reference might then be created within which barely perceptible length differences could acquire meaning.

To decide definitively between these hypotheses a detailed exploration would be necessary of the minimum length difference discriminated as a function of age, both with and without measurement, and with stimuli which do not give rise to an illusion. It would be interesting to know whether the smallest length difference noticed by children decreases steadily as a function of age, or whether there is a sharp decrease at any period which might correspond to the change in "attitude" toward small differences which is suggested by the present data.[18] If there is such a sharp decrease, it would be interesting to know whether or not it is associated with the development of inferential responses of the kind studied here, and whether these inferential responses usher in the sharp decrease in DL, or follow it. Some of the issues raised by this study might then be resolved.

Recapitulation

Some aspects of the logic of length measurement were discussed and a series of studies of length judgment were reviewed. These included a group of studies by Piaget in which he attempted to diagnose children's grasp of the logic of measurement through their manner of using measuring rods. He found children under 7 yr. deficient in this respect. A number of criticisms of Piaget's work and general methodology were discussed. It was suggested that adequate use of measuring rods by children may involve factors other than the performance of the logical operations basic to measurement. In particular, learning of effective measuring techniques may be an important factor.

To investigate this possibility a task was constructed in which no manipulation of measuring rods was demanded of the *S*s. They were required to make deductions about the relative lengths of two nearly equal upright sticks, using the rule, $A > B . B > C . \supset . A > C$. This inference has the status of an axiom in the logic of measurement. The experimental procedures were adapted from work with animals, so that it was possible to present the task with a minimum of verbal communication between *E* and *S*. Vocabulary development, an uncontrolled factor in Piaget's studies, is therefore unlikely to be a factor influencing the results. The methodological importance of this innovation was stressed.

---

[18] It is assumed, of course, that the DL is explored using an experimental technique in which *S* is first trained to discriminate large length differences, and then the smallest length difference to which he will generalize this discrimination is studied.

It was found that with these procedures, the inferential response $A > B . B > C . \supset . A > C$ was elicited approximately two or three years before the age at which Piaget claims it first becomes available to children.

Unless Piaget incorrectly reports the age at which children begin to manipulate measuring rods properly, it is clear that factors other than correct reasoning according to the axiomatic standards of the logic of measurement are involved in his studies. These factors are probably vocabulary development (particularly, comprehension of the meanings of words such as "measure," "same length," etc.) and the development of skill and interest in the technique of length measurement. The effect of these factors is probably to conceal the reasoning ability of many of Piaget's *S*s.

The results indicate that the definition of "emergent levels" of development is a complicated problem. The objection that the apparent "age of emergence" is dependent on the specific experimental procedures utilized does not pose insuperable difficulties for the concept of "levels." This objection does, however, suffice to demonstrate that any stated "age of emergence" must be arbitrary until the effects of all relevant variables have been systematically explored, so that extraneous factors (e.g., vocabulary development, skill in the use of measuring rods, etc.) which may hinder the younger *S*s from responding to the relevant cues (the relative lengths of uprights and measuring stick) are identified and eliminated. With this qualification, the experiment reported lends support to Piaget's notion that the formal reasoning inherent in length measurement constitutes an "emergent" process in child development.

The experimental task also contained some trials in which the two nearly equal uprights were compared by direct visual comparison. Qualitative aspects of the behavior of some of the older *S*s in these trials suggested that they adopted an "attitude" or "set" which was different from that of many of the younger *S*s, and which assisted them to achieve some measure of success in these discrimination trials. The possible genetic relationship of this change with age in "attitude" or "set" to the development of the inferential response was discussed. The evidence, although not conclusive, suggests that once a child is capable of the inferences basic to measurement, he may develop a concept of length as a dimension, as a result of which he more readily adopts a "set" to discriminate small differences in length. A method of examining this hypothesis experimentally was suggested.

DEVELOPMENT OF A CONCEPT OF POSITION ORDER

In this section Piaget's views on the development of concepts of order are examined in a way which parallels the previous examination of his views on the development of length measurement. A study of his is described and an experiment is reported which investigates certain task parameters unexplored by Piaget.

Piaget's Study

In the Piaget and Inhelder (1956, Ch. 3) study, seven to nine beads of different colors threaded on a wire served as a model which *S* had to copy by choosing similar colored beads and threading them in the same order on another wire. For some of the *S*s, instead of beads on a wire, dolls' garments differing in color or form or both were hung on a toy clothes line, and *S* was told to choose similar garments and to hang them in the same order on another clothes line. The *S*s' ability to distinguish among the beads and garments was established before the experiment.

Piaget required his *S*s to copy the order of a linear model and also to make various transformations, as demanded by the following spatial arrangements of model and copy: 1. model linear, copy linear and directly beneath the model; 2. model linear, copy linear but displaced to one side; 3. model circular, copy linear; 4. model linear, copy linear with order inversed; 5. model in the form of a "figure 8," copy linear.

Piaget does not discuss how he conveyed the instructions to the *S*s. This must have been a problem since *S* need not interpret the word "same" to mean "same order," and any instruction which contains the word "order" presumably will only be understood by a child who already possesses a concept of order to which he can assimilate this word. Neither does Piaget mention how he specified the starting position to the child when the model was nonlinear.

Piaget does not publish any tabulated results. He says the responses could be divided into three stages (with substages) of development between 3 and 7 yr. In Stage I, from about 3 to 4½ yr. of age, *S* was able to select similar beads or garments but was not able to reproduce accurately the order prevailing in the model. In Stage IIA, the model was correctly reproduced only when linear and directly above the copy, so that careful matching of each bead with the one immediately above it was possible. In Stage IIB, the copy was correct when displaced to one side of a linear model, and also when the model was circular, but failure still occurred on the other two arrangements, indicating to Piaget that the notion of position order was not yet completely abstracted. In Stage III, beginning at about age 7 yr., all arrangements were accurately copied.

Whether the material to be ordered consisted of colored beads or of dolls' garments was apparently immaterial.

It may be noted that Piaget takes it for granted that the five settings he used are on a continuum of difficulty. Placing them all within the same logical category of transformations of ordered sequences, he assumes that the five settings represent stages of development of the same psychological process. However, it is possible that there might be psychological processes or abilities involved in transforming a cyclic order or "figure 8" into a linear order, which are not involved in a linear-to-linear translation. This could only be investigated through an examination of the intercorrelations of success and failure between the settings (not reported by Piaget), or by a longitudinal study of the development of individual children through the stages.

In some of the omnibus intelligence tests there are individual items in which success appears to demand some concept of order. For example, on Form M of the Binet there is an item which demands copying a chain of beads, which remains exposed. Test standardization (Terman & Merrill, 1937) places this test at Year VI. No disagreement with Piaget's findings is suggested.

In tasks which require the copying or matching of sequences, such as those described, there are three obvious parameters. The formal arrangement of the stimulus (linear order, inverse linear, cyclic, etc.) is one such parameter. This is the one varied by Piaget.

The number of objects to be ordered is another obvious parameter. One might reasonably expect that a sequence of three objects would prove less difficult than sequences of five or seven or more. This parameter is studied in detail in the experiment reported below.

The third obvious parameter is the nature of the material ordered. Although Piaget found that copying a sequence of dolls' garments did not differ in difficulty from copying a sequence of differently colored beads, this parameter has not been studied systematically. Two possible content parameters are the degree of familiarity of the material and its attractiveness to the child. Examination of content parameters is important, since if order concepts are entirely content-specific, i.e., if as many concepts of order develop in children as there are objects capable of being arranged in sequence, then one would have to question the usefulness of Piaget's postulation of the development of a general concept of position order.

In addition to the above parameters, there may be other factors involved. It is possible to devise a number of different experimental procedures for eliciting order concepts, and it is likely that the age at which order concepts will be found to develop will vary with the specific experimental procedures selected. In Piaget's study the instructions were communicated verbally and the child's response was to copy a model provided by *E*. The adequacy of the verbal communication between *E* and *S* therefore becomes a factor of unknown importance.

In a single experiment it is impossible to study the effects of varying all the above parameters. In the study reported below the parameter "number of objects ordered" has been singled out for special investigation. While recognizing that content parameters may later prove important, it seemed that the parameter selected should be studied first, and the material remained the same throughout the experiment. It consisted of colored discs pasted on cards and only the number of colors was varied. The experimental procedures were developed from nonverbal techniques used in work with animals. The "matching-from-sample" technique was employed. There were few verbal instructions, and the words "same" and "order" were eliminated. This use of relatively nonverbal methods constitutes an important departure from Piaget's experimental technique.[19] All stimulus settings were linear; consequently the ages at which order concepts were elicited are not necessarily generalizable to stimulus settings which are nonlinear, like some of those used by Piaget.

Method

Subjects

The *S*s were 46 children from the nurseries described in the report of the experiment in the previous section. They ranged in age from 3-6 to 7-3. The distribution by age is shown in Table 3. There were 20 boys and 26 girls. Thirty-seven *S*s also received the measurement task described earlier, and of these, 21 took the measurement task before being given the order-discrimination task.

Materials and Procedure

The task consisted of a series of problems in which *S* was required to find an item identical to a sample item. The task began with simple color matching: *S* had to find that one of three discs which was the same color as a sample disc. In subsequent problems *S* had to match color sequences. In the initial series of these the sequence was only two units long, e.g., *S* had to match a sequence in the order red-gray (say) against one of two sequences, gray-red and red-gray. In the next series of problems the sequence had three units, e.g., *S* matched a sample in the order red-gray-green (say) against one of three sequences, gray-red-green, red-green-gray, and red-gray-green. In the fourth and fifth series of problems the sequences were five and seven units long, respectively.

Materials. These consisted of blank filing cards on which colored paper discs 1 in. in diameter had been pasted. The colors used were red, yellow, green, black, walnut, Prussian blue, and a light sky-blue gray. There were five sets of cards. Set A consisted of 14 cards with a single disc in the center of each. Set B consisted of 21 cards, with a pair of unlike colors on each card. Seven pairs of colors were selected and each pair was represented three times. Set C consisted of three sets of six cards with three colors on each card. In Set D there were two sets of six cards with five colors on each card. Set E comprised 10 cards with all seven colors on each card.

In each of Sets B, C, D, and E, the sequence of colors on any card could be reversed by inverting the card. The precise sequences of the colors on the cards in each trial is shown in Table 2. It can be seen that in Sets D and E the first and last colors were the same on all cards in each trial so that *S* could not match the sample without paying attention to the internal order of the colors.

Procedure. The *S* sat facing the short edge of a low rectangular table and *E* sat on his right facing the long edge. The sample card was placed flat on the table immediately in front of *S* and the choice cards (i.e., the cards from which *S* was to choose the one matching the sample) were placed equidistant from him about 2½ ft. away on the table, where they could easily be manipulated by *E*. Placing the choice cards beyond *S*'s convenient reach was found to be the best procedure. Being farther away, the choice cards were closer together in his field of view, thus encouraging comparison behavior instead of an unthinking dive for one of the cards. This arrangement also prevented *S* from attempting to reach for more than one card at once, thus removing a possible source of discord between *E* and *S!*

The choice cards rested against small white open cardboard boxes placed face down on the table. The cards were thus supported at an angle 30°–45° with the table surface. A small piece of candy was placed under the box which supported the card which duplicated the sample. Between trials a cardboard screen placed on the table hid the boxes from *S*, and *E* was careful that no noise or other cues occurred as the candy was placed in position. There were always three choice

---

[19] See the discussion of this point in the previous section.

TABLE 2

Color Sequences Used in Series B, C, D, and E of the Order-Discrimination Task

Trial	Sample	Choice Cards
	Series B	
1–6	RS	RS and SR
7–12	SR	RS and SR
13	PY	PY and YP
14	YP	PY and YP
15–26	as Trials 13–14, using BS and SB, GW and WG, etc.	
	Series C	
1	SGR	SGR, GSR, and SRG
2	GSR	SGR, GSR, and SRG
3	SRG	SGR, GSR, and SRG
4–6	as Trials 1–3, using R, B, and Y	
7–9	as Trials 1–3, using Y, W, and P	
10–18	as Trials 1–9, cards inverted	
	Series D	
1	RPYGW	RPYGW, RPGYW, and RYPGW
2	RPGYW	RPYGW, RPGYW, and RYPGW
3	RYPGW	RPYGW, RPGYW, and RYPGW
4–6	as Trials 1–3, using Y, R, B, S, and G	
7–12	as Trials 1–6, cards inverted	
	Series E	
1	SWBGYRP	SWBGYRP, SBWGYRP, and SWBGRYP
2	SBWGYRP	SWBGYRP, SBWGYRP, and SWBGRYP
3	SWBGRYP	SWBGYRP, SBWGYRP, and SWBGRYP
4	SWBYGRP	SWBYGRP, SWGBYRP, and SWBGYRP
5	SWGBYRP	SWBYGRP, SWGBYRP, and SWBGYRP
6–8	as Trials 1–3, cards inverted	

Note.—B = black, G = green, R = red, Y = yellow, W = walnut, S = sky-blue gray, P = Prussian blue.

cards, except when the cards of Set B were used, when there were only two. The position of the correct choice card was varied randomly, with the restriction that the same position was never correct in more than two successive trials. As the sample was presented to *S*, *E* asked, "Where is this one?" or said, "Now find this card." *S* indicated his response by pointing and was given the candy if correct. If the pointing was equivocal, the boxes were moved just within *S*'s reach so that he could find the candy himself.

The following "training" procedure was adopted in all trials, except in the first 12 trials of Series B. If *S*'s response was incorrect, he was first shown that there was no candy under the box to which he was pointing. Then the sample was held over the correct card with the words, "Look, here it is," and the box was lifted to expose the candy which, however, was not given to *S*. When *S*'s response was correct, the sample was held over the indicated card and *S* was told, "That's right." The box was lifted and the candy given to *S*.

Five series of trials were given, each series employing the cards of one of the sets. With Series A which was presented first, each *S* was taught to match from sample. This series was discontinued following six consecutive correct matchings. Series B began with six trials in which the sample was a sequence red-gray, and the choice cards were gray-red and red-gray. In the first four of these trials, after the sample had rested in front of *S* for a few seconds, it was picked up and held over each of the choice cards in turn for about 10 sec. to facilitate visual comparison of the sample and choice cards. The *S* was asked, "Is it this one? Or is it this one?" Any pointing by *S* during this period was ignored. The last two of the six trials followed the regular procedure. These six trials using the sample sequence red-gray were followed by a similar six trials with the sample gray-red, and the same choice cards. As shown in Table 2, there then ensued 14 generalization trials in which each sequence in Set B served as sample once. The *S* was considered successful in Series B if he met either of the following criteria: (*a*) correct matching in all of the first five generalization trials, in which case Series B was discontinued, or (*b*) correct matching in 11 of the 14 generalization trials. All *S*s who failed to meet either of these criteria went through all 14 trials.

Series C was administered whether or not the criteria for success on Series B had been met. Series C consisted of 18 trials, all of which were treated as generalization trials. These trials are shown in Table 2. The *S* was considered successful in Series C either following five consecutive matchings, or if correct matching occurred in 11 of the 18 trials.

Series D was administered unless *S* had failed to meet any of the criteria for success in either Series B or C.[20] Series D consisted of 12 trials, as shown in Table 2. Criteria for success were either five consecutive correct responses, or a total of 8 out of 12 trials correct. If either of these criteria was met, Series E, with seven colors per card, was administered. This series consisted of eight trials with passing criterion six trials correct.

The criteria set for success in the five series were selected so that one would expect them to be attained by approximately 5% of *S*s who responded randomly.

Administration of the entire order-discrimination task often required more than one session, although rarely more than two sessions. At the beginning of all sessions subsequent to the first, Series A was re-administered.

Results and Discussion

Level of performance attained. With the exception of two boys who were therefore excluded from the remaining four series, all 46 *S*s readily learned to match from sample when there was only one color on each card (Series A). Twenty-two *S*s met one or another of the criteria for success on Series B and 22 failed; on Series C, 29 *S*s met a criterion, 15 failing. The 22 *S*s who succeeded on Series B also did so on Series C. With Series D there were 25 successes and with Series E 11 successes.

The development with age of success on each series is shown in Table 3. It can be seen that no child under 4-2 showed any concept of order at all. Surprisingly, significantly fewer children matched the order of sequences of two items than the order of sequences of three items.[21] Order was apparently most readily discriminated when there were three colors on the cards.

One possible explanation for the differing incidence of successful matching on Series B and C was examined. It was thought that in responding to Series C, the children might have profited from the training provided by the trials of Series B, even though they failed to match on these trials. This explanation implied that if some of the trials of Series B were repeated following the administration of the later series, matching of order would be found on Series B. To explore this possibility, four of the seven *S*s who failed Series B but solved Series C were given between 5 and 10 extra trials on Series B again after the longer series had been administered. All four of these *S*s made several errors, indicating that

TABLE 3

PERCENTAGE *S*S IN EACH AGE GROUP REACHING CRITERIA FOR SUCCESS IN THE SERIES OF THE ORDER-DISCRIMINATION TASK

Age Group	*N*	No. of Colors in Sequence			
		2	3	5	7
3–6 to 4–2	7[a]	0	0	0	0
4–3 to 4–11	12[a]	25	50	33	0
5–0 to 5–8	15	67	87	73	27
5–9 to 6–5	6	83	100	100	67
6–6 to 7–3	4	100	100	100	75

[a] One *S*, who failed to learn to match-from-sample with one color per card, is excluded.

---

[20] It seemed unlikely that a child who had failed to discriminate order with either two or three colors per card would suddenly begin to succeed when cards containing five or seven colors were introduced. Moreover, while the 12 trials of Series D were sufficient to demonstrate generalization of order discrimination from two or three colors per card to five colors, it would be quite unfair to expect *S*s who had not demonstrated order discrimination in the 44 trials of the previous two series suddenly to demonstrate this within the mere 12 trials of Series D.

[21] Chi square (correlated proportions, $df = 1$) = 7.1 ($p < .01$). Sequences of two and five items did not differ significantly in the number of *S*s solving them.

practice with the longer series had not caused any striking improvement in performance on Series B.

This seemed to leave no good explanation for the greater difficulty of Series B over Series C. The children's verbalizations provided no clues; there were in fact practically no task-related verbalizations during the entire task. No particular differences in eye movements were noticed during responses to Series B and C. Latency of response was not noticeably greater on Series B than on Series C, although in general children who matched successfully appeared to take somewhat longer to respond on all series. An interesting possibility[22] is that Series B may have been more difficult because the order of the colors on the wrong choice card is a symmetrical reversal of the sample sequence on this series, whereas in none of the subsequent series do the choice cards contain an inversion of the sample.

There is no evidence that success in discriminating order is affected either by *S*'s sex or by prior experience with the measurement task. Of 18 boys, 56%, 61%, 56%, and 22% succeeded on Series B, C, D, and E, respectively. The corresponding figures for the 26 girls were 46%, 69%, 58%, and 30%. The differences between boys and girls are all statistically insignificant.

Of the 44 *S*s, 21 had prior experience with the measurement task. Of these, nine failed to show any concept of order. Of the 23 *S*s without prior experience with the measurement task, six failed to show any concept of order. This difference is not statistically significant. The paradoxical direction of the difference may be accounted for by the fact that, as explained earlier, a majority of the younger *S*s were given the measurement task first.

Significance of results for Piaget's theory. In his experiment reviewed above, Piaget used sequences which contained at least seven items. The arrangement used by him which most nearly corresponded to that used in this experiment is the one with a linear model and with the copy displaced to one side. Piaget placed his children who succeeded in copying the bead order in this arrangement in Stage IIB[23] which ranges from about 5½ to 7 yr. The results reported here are consistent with Piaget's empirical findings, since the lowest age group which showed a large measure of success in matching the order of linear sequences of seven items was aged 5-9 to 6-5.

While confirming Piaget's empirical findings in this particular, the results also show (Table 3) that sequences of seven items are substantially more difficult to match than shorter sequences. Those *S*s as young as 4-3 to 4-11 showed a considerable degree of success in ordering three items. It is possible that the greater difficulty of the sequence of seven was due, not to any lack of understanding of order per se, but to the greater demands made by the longer sequence upon the *S*s' attention, interest, and persistence, or to other factors of a similar nature. This is consistent with the *S*s' behavior with the sequence of seven. The older *S*s took longer to compare the sample with the choices available and were more thorough in their comparisons, often carefully matching the colors one by one and indicating them on the sample with their fingers.

Piaget's interpretation of his results is that a concept of order does not mature fully in children until around age 7 yr. The results of this experiment suggest that whatever logical ability is involved in these tasks, it is already present in many children before age five, and the possibility of a yet earlier development is not excluded.

The present experiment differed from Piaget's, not only in that the number of items in linear sequence was varied, but also in the use of relatively nonverbal techniques. Where the sequence contained seven items, the fact that Piaget's results were comparable to those reported here, suggests that for these sequences it made little difference whether Piaget's verbal instructions or the relatively nonverbal matching-from-sample technique was used. This equivalence may well have been because the *S*s who solved the sequences of seven items were the older children who were probably better acquainted with the words used in the instructions. Since Piaget did not use sequences of less than seven items, it is unclear to what extent the nonverbal technique facilitated the solutions of the younger *S*s on these sequences.

---

[22] Suggested by Brewster Smith.

[23] Stage IIB is also defined by success with the circular-to-linear translation. The present study, which investigates only linear-to-linear transformations, neither corroborates nor disproves Piaget's views of the development of nonlinear order concepts. It has already been pointed out that Piaget gives no direct evidence which demonstrates that the same psychological process is involved in linear and nonlinear transformations.

The important effect of the use of the nonverbal method is that lack of language comprehension cannot be held responsible for the total failure of some *S*s to discriminate order on any of the series.

The criticism of Piaget which is suggested by the data is similar to the criticism made in the preceding chapter of his views on the development of the formal reasoning inherent in measurement. In his writings Piaget shows little sign of appreciating the fact that the age at which the responses he studies can be elicited is profoundly dependent on the specific experimental conditions he employs. When the ease with which a response can be elicited is greatly altered by varying certain task parameters, is it legitimate or profitable to attempt to state an age at which the response develops? In the case of order discrimination, might one not conclude that it is unsatisfactory to postulate a logical reasoning process which defines an "emergent level" of development, since one always has to consider how many objects are involved, perhaps also the nature of the objects, and how the task is presented?

While this line of attack is certainly not fatal for the notion of levels of development, it does demonstrate the necessity for systematic exploration of the factors that inhibit or facilitate solution of order-discrimination tasks. Obviously, if one seeks to state an age at which an order-discrimination response develops, the only age which is not completely arbitrary is the earliest age at which order discrimination can be unequivocally elicited using the simplest experimental procedures. One cannot know the simplest procedures until one has systematically explored all parameters which are not involved in the definition of order discrimination. The experiment reported is a contribution in this direction.

Systematic parametric studies will lead to one of the two following results:

First, it may be found that a point is reached at which no further simplification of procedure is possible, without so profoundly altering the nature of the task that "order" ceases to be the relevant cue. This point will presumably be reached when the extraneous factors whose intervention prevents the youngest *S*s from solving the task have been discovered and eliminated.

Second, it may be found that the concept of order discrimination disintegrates on analysis. That is, it may prove impossible to define "sequence" as a cue in terms of clear-cut experimental operations. There is nothing in the results to suggest that this will be the case and Piaget's concept of levels of development is therefore broadly confirmed. However, a conclusion of this nature should be advanced with caution, since there are many instances in the history of comparative psychology where distinctions between "levels" have been found untenable when analyzed.[24]

---

[24] An instructive example is the distinction once made by Bingham (1914, 1922) and Hunter (1913) between perception of "form, per se" and perception of "shape" or "pattern." When trained to discriminate two forms, "lower" organisms were believed to have learned to respond to a particular retinal distribution of light ("pattern" vision). "Higher" organisms, who freely generalized whatever they had learned to many stimuli possessing similar geometric properties to those on which they were trained, were said to be responding to "form per se." However, analysis demonstrated the unreality of this distinction. It was found that "generalization after learned discrimination between two stimuli is an obvious and universal phenomenon" (Andrew & Harlow, 1942, p. 18), and that response is never simply to a particular retinal distribution of light. Lashley concluded that there is "no evidence that pattern or shape are ever the basis of reaction . . . neither phylogenetic studies nor evidence from pathology reveals pattern vision as a primitive function" (Lashley, 1938, p. 138).

It is clear that parallel issues are raised by the results of the experiments described in this and in the preceding section. That is, of course, no accident: it is likely that the same issues will present themselves whenever any of the formal reasoning processes studied by Piaget are the subject of experiment. The theoretical question is whether the kinds of reasoning whose formal structure is described in logic and mathematics constitute emergent psychological processes or whether, when the experimental procedures used to elicit these kinds of reasoning are systematically varied, it is found that the distinctions between the "emergent" and other more "primitive" processes are not translatable into experimental terms so that clearly definable levels emerge.

Recapitulation

A study by Piaget which explores the development of children's concepts of position order was discussed. It was pointed out that order discrimination is probably a function of several variables unexplored by Piaget—for example, the number and nature of the objects in sequence, and vocabulary development when the test instructions employ the word "order," or synonyms.

An experiment was designed which employed the matching-from-sample procedure developed in work with animals, so that the results are unlikely to be affected by vocabulary development. The number of objects in the sequences to be matched was varied systematically from two to seven, so that the development of order-discrimination responses in children could be explored as a function of this variable.

It was found that with those sequences which contained approximately the same number of objects as the sequences used by Piaget (i.e., seven objects), results comparable to his were obtained. However, it was also found that when the number of objects in sequence was reduced, order discrimination was elicited at an age considerably earlier than the age at which Piaget claims it develops in children. The importance of the variable, number of objects in sequence, was therefore demonstrated. It is likely, though not proven, that correct responses among the younger *S*s were facilitated by the relatively nonverbal presentation of the task.

Piaget's concept of "levels" of development was discussed. The evidence indicates that in order to establish discrimination of sequence as an "emergent" process, the effects of varying a number of factors other than the "order" cue itself require exploration. However, there is nothing in the results to suggest that the relational cue of spatial sequence cannot be given clear experimental definition and the data support the view that response to this cue may constitute an "emergent." To this important extent Piaget's concept of "levels" is confirmed.

The Relation of Order Discrimination to the Inferential Response $A > B . B > C . \supset . A > C$

In the previous sections it was shown that, although Piaget may be mistaken about the specific ages at which order discrimination and the inferential response $A > B . B > C . \supset . A > C$ develop, there is nothing to indicate that his concept of levels of development should be rejected.

In this section and the following one, Piaget's theory of intelligence, which contains his definition of general levels of intellectual functioning, is discussed. This theory suggests specific predictions about the relationship between discrimination of order and the inferential response studied.

Piaget's Theory of Intelligence

According to Piaget, logical reasoning consists of a group of "emergent" intellectual functions. He refers to these as "operational thinking," with which he identifies "intelligence." His theory of intelligence, therefore, attempts to give a detailed definition of logical reasoning in the following manner.

Piaget's basic postulate is that mature intellectual processes are isomorphic with the "operations" of bivalent logic.[25] An example may clarify this use of the term "isomorphic." "Addition" of numbers is an intellectual process occurring in children when they do arithmetic; it is presumed to be isomorphic with the "additive operation" discussed by mathematicians. Similarly, deduction is an intellectual process (or processes) which can be placed in correspondence with one or more logical definitions. Piaget's definitions of the logical operations with which intellectual processes are assumed to be isomorphic are similar, but not identical, to the operations of the spatial, class, and propositional interpretations of the Boolean calculus, as described in texts on logic. Thus, the principal processes for which he finds isomorphic logical definitions are conjunction and disjunction of classes, of regions of space and of propositions,[26] and the corresponding relations of class inclusion, geographical or spatial inclusion, and implication. These processes are assumed to be, so to speak, the "atoms" of which intelligence is made. This part of Piaget's theory is essentially an essay in formal logic rather than psychology. According to Piaget, the relation between the logical definitions and the corresponding intellectual processes is analogous to the relation between pure geometry and physical geometry, or to the isomorphism between complex algebra and certain properties of alternating current electric circuits.[27]

Psychologically, the terms *conjunction* and *disjunction* are ambiguous. To illustrate the meaning intended here, class conjunction will be discussed in some detail. A conjunction of classes is that class whose members share the attributes of two other classes: thus, the class of white sheep is the conjunction of the class of white objects and the class of sheep, the class of balls (with exceptions) is the conjunction of the class of spheres and the class of playthings.

It might be argued that whenever we identify an object we perceive (e.g., a white sheep) we are performing a conjunction of classes, since any perceptual identification requires the simultaneous recognition of a number of different cues. If class conjunction is interpreted psychologically as the recognition of combinations of cues, it is obvious that it is an extremely primitive process which occurs early in phylogenetic development: many nonverbalizing organisms certainly discriminate classes of objects on the basis of combinations of cues. The apparent ubiquity of logical operations, viewed from this standpoint, appears to have led Lashley (1956) to assert that the operations of bivalent logic are coincident with the existence of nervous tissue, and this has led him to insist upon the need for a "neurology of logic."

Whether or not one selects a logical model for perception appears to be a matter of taste; many workers prefer a statistical model or one taken from information theory (e.g., Attneave, 1954; Binder, 1955). It is clear that whatever the reasoning processes are to which Piaget refers in his studies, they are *not* the processes involved in perceptual identification as discussed in the preceding paragraph. Unfortunately, the writer has not been able to find an adequate discussion by Piaget of this ambiguity in his concept of logical operation;

---

[25] It is difficult to phrase this postulate so as to avoid becoming embroiled in epistomological argument. The epistomology of the connection between reasoning process and logical operation is still disputed among logicians. To identify logical operation and reasoning process is Boole's fallacy and Piaget is well aware of this. The term "isomorphic" has been selected to express the postulated relation between logical operation and intellectual process in order—hopefully—to circumvent the epistomological issues. The use of this term does not do violence to Piaget's generally expressed views (Piaget, 1949, pp. 1–31; Piaget, 1950, pp. 18–37; Piaget, 1953b).

[26] The logical definitions of these operations may be found in most introductory texts in symbolic logic (e.g., Langer, 1953). Piaget's own definitions are expressed somewhat differently, but a review of these differences would complicate this discussion unnecessarily. See Piaget (1949; or 1950, pp. 43–47) for a brief summary of Piaget's system. For examples of these operations in ordinary speech, see Footnote 3.

[27] In a book review, Isaacs (1951) discusses this point in relation to Piaget's theory.

consequently, the interpretation of his conception that follows is the writer's.

In the English language, a conjunction of classes is denoted by a syntactical device: the qualification of a noun by an adjective as, for example, in the phrases "white sheep," "navigable rivers," etc. To render the following discussion more concrete, suppose that the English language be modified so that a conjunction of classes is expressed by the nonsense syllable "ZIK." In this modified English, white sheep would be denoted by the phrase "ZIK-whitethings-sheep"; navigable rivers would become "ZIK-navigablewaters-rivers"; balls would be denoted by the same word, "balls," but might also be represented by the phrase "ZIK-spheres-playthings." It is clear that the words "balls" and "ZIK" have quite different meanings. Knowledge of the meaning of the words "ball" and "white sheep," reflected in correct identification of these objects, is not the same as knowledge of the meaning of the word "ZIK," reflected in the proper usage of this term. The latter, let us say, reflects a general concept of class conjunction, which is not indicated by perceptual discrimination of either balls or white sheep. I believe it is to this, the discrimination of the meaning of the word "ZIK," that Piaget's theory refers when he uses the term *multiplication* or *conjunction of classes.*

There are many words and phrases in the English language which have a meaning as abstract as that of "ZIK" in the modified English discussed above. "Like," "similar to," "same as," "analogous to," "type of," "kind of," "species," "genus," "both . . . and," "either . . . or," "is a," "common to," are a few examples. These words express general characteristics of classes and have no reference to any specific class. The intellectual processes which Piaget's theory, as it is construed here, treats as isomorphic to the logical definitions of conjunction and disjunction of classes and of the relation of class inclusion are the kinds of processes which are reflected in the understanding and proper usage of such words and phrases.

The distinction made between the learning of names of objects and the learning of words such as "ZIK" is not new to psychology. It seems closely related to the distinction Lashley (1938) has made between "first-order generalizations" and "second-order generalizations." Examples in the literature of first-order generalizations are learning to discriminate "triangularity per se" (Andrew & Harlow, 1942; Fields: 1932, 1936; Gellerman, 1933; Lashley, 1938),[28] and to distinguish the identical elements in Hull's (1920) Chinese ideographs. Lashley's example of a problem whose solution requires second-order generalization is the "oddity" problem (Harlow, 1951; Lashley, 1938). The writer would suggest that other examples among nonverbal tests are the "sameness-difference" problem administered to chimpanzees (Robinson, 1955), and the experiments reported in the present paper. Examples of verbally communicated problems involving second-order generalization are provided in studies of the learning of logical relations by children (Kreezer & Dallenbach, 1929; Schooley & Hartmann, 1937).

It may be noted that the distinction between "abstract" and "concrete" thinking made by Goldstein and Scheerer (1941) appears to have much in common with Piaget's distinction between "operational" and "pre-operational" thinking and with Lashley's between first- and second-order generalizations.[29] All these distinctions appear to follow that made by logicians between "class" and "class of classes."[30]

Assuming the validity of the postulate that mature intellectual processes are isomorphic with the operations of bivalent logic,[31] Piaget adopts the following research method. He sets children problems whose solution requires a certain line of logical reasoning. He then observes how children at various ages solve these problems, and he tries to observe the method of solution. Consideration of the ages of the children who solve the problem leads Piaget to specify the age at which the reasoning process involved in the problem develops.

---

[28] That is, learning to choose triangular stimuli, as against other shapes, irrespective of the size, color, or orientation of the stimuli.

[29] However, as in his discussion of amnesic aphasia, Goldstein (1948) sometimes appears to suggest that it is the nominal aspects of speech which are most clearly related to "abstract" thinking, whereas in Piaget's theory, it would be, of course, the syntactical aspects.

[30] Cf. Russell's (1920, Ch. 2) definition of number.

He finds that certain types of problems are solved at the same age. Although he never explores the intercorrelations between success and failure on different problems, he concludes that reasoning processes develop in groups at particular ages. Specifically, he claims that the processes of class conjunction, class disjunction, and the concept of the relation of class inclusion develop in children at approximately 7 yr. of age. He also claims that the analogous operations of the spatial interpretation of Boole's system[32] develop at the same age. At around 11 yr., the child becomes capable of reasoning with propositions. That is, the processes of conjunction and disjunction of propositions develop at this age along with the concept of the relation of implication.

It can be seen, therefore, that Piaget's notion that reasoning processes develop in groups clearly implies the postulate that where operations, inferences, etc., are mutually interdependent (i.e., the logical definition of the one involves reference to the other, or that both derive from the same interpretation of a logical calculus), then the corresponding reasoning processes develop in association in children's thinking.[33] Thus, the operations of class conjunction, class disjunction, and the concept of the relation of class inclusion all derive from the class interpretation of the Boolean calculus and would therefore be expected to develop in children as a group. Later it will be shown that position order and the proposition $A > B . B > C . \supset . A > C$ (where "$>$" is interpreted "longer") are logically interdependent and, as an illustration of how Piaget's theory might operate, this postulate will be used to predict a high correlation between success in the tasks described in the two previous sections.

Summarizing Piaget's theory schematically, it consists, first, of a set of logical definitions with which the intellectual processes subsumed under the title "intelligence" are postulated to be isomorphic. Second, it postulates implicitly that where the logical definitions are interdependent or derive from the same interpretation of a logical system, the corresponding intellectual processes will be found to develop in association in children's thinking. Third, the theory contains a group of empirical laws stating the ages at which certain groups of processes develop. One may note that the ages indicated in these empirical laws could readily be altered where necessary without changing the main structure of Piaget's theory.

A criticism of Piaget's concept of intellectual process. In theory construction in psychology, intellectual processes have the theoretical status of hypothetical constructs. A hypothetical construct requires two kinds of specification. First, it is specified in terms of its relations to other constructs of a theoretical system. In Piaget's theory this kind of definition is contained in the two postulates that intellectual processes are isomorphic to definitions of logical operations, and that when the logical definitions are interdependent the corresponding intellectual processes develop in association. However, in modern behavioral theory a hypothetical construct of this kind requires specification of the experimental operations through which it may be elicited, and of the characteristics of the response which testifies to its presence; without this it lacks empirical reference. Unfortunately, there is little in Piaget's theory which fulfills this function. The goal of his experiments is not to provide adequate specifiying procedures, but only to provide situations in which he can diagnose the character of intellectual processes through a series of acute observations of the *S*s' behavior and language. The tasks he uses are therefore not intended to serve as specifying experimental procedures, and the results from the experiments reported in the previous sections demonstrate that the tasks used by him in his studies of length measurement and of concepts of position order are inadequate for this purpose. It may be that Piaget's view that psychological processes do not need to be specified in terms of experimental procedures leads him to neglect the need for careful experimental design, with the result that he sometimes overlooks important variables.

---

[31] One may note that there is no a priori reason why bivalent logic, rather than any of the polyvalent logics, should be selected as the model for reasoning.

[32] As noted in Footnote 3, Piaget's system differs somewhat from Boole's, but the differences can be neglected for the purpose of this study.

[33] The proposition that logical and ontogenetic priority correspond would seem to be a corollary. That is, where one logical definition is dependent on another, the intellectual process corresponding to the latter develops earlier. That there is at least some truth in this proposition is suggested by many educational practices based on common sense: for example, children are customarily taught to add numbers before they are taught to multiply, presumably because multiplication logically presupposes addition.

This criticism exposes an important weakness of Piaget's theory—namely, that its roots in empirical fact are not entirely adequate. The power of his theory lies in the fact that it is potentially capable of giving a detailed and precise account of the development of intellectual functioning, and that numerous predictions may be derived from a few simple postulates.

Derivation of the Genetic Association of Order Discrimination With the Inferential Response $A > B . B > C . \supset . A > C$

As an example of how Piaget's theory might operate, this relationship will be derived using the postulate that when definition of logical operations are interdependent, the intellectual processes isomorphic to these operations develop in association. It follows from this postulate that, if there is a clear relation between the definition in logic of serial order and the proposition $A > B . B > C . \supset . A > C$, then discrimination of position order and the inferential response $A > B . B > C . \supset . A > C$ develop together in children.

The relation in logic between "order" and rules of the form $A > B . B > C . \supset . A > C$ is clearly stated by Russell in a discussion of the definition of order. Russell (1920) states, "We must not look for the definition of order in the nature of the set of terms to be ordered. . . . The order lies, not in the *class* of terms, but in a relation among the members of the class, in respect of which some appear as earlier and some as later" (pp. 30–31).

According to Russell, ordering relations are asymmetric, transitive, and connected.[34] That is, the property of transitivity of a relation is part of the definition of order, or, as Nagel (1930) states, it is an "axiom" of order. The rule, $A > B . B > C . \supset . A > C$ merely states the transitive nature of the relation "longer than."

Given that a relation is asymmetric,[35] and disregarding the property of connectedness,[36] one may therefore predict that as the relation becomes transitive for a child, he should begin to distinguish different sequential arrangements of objects ordered with respect to it.

---

[34] A relation is asymmetric if, holding between *x* and *y*, it does not also hold between *y* and *x* (e.g., longer than, left of, father of). A relation is transitive if, holding between *x* and *y*, and also between *y* and *z*, it holds between *x* and *z* (e.g., longer than, left of, same as). A relation is connected if it has the property that, "given any two terms of the class to be ordered, there [is] one which precedes and another which follows. For example, of any two integers, . . . one is smaller and the other greater; but of any two complex numbers this is not true" (Russell, 1920, p. 32).

[35] I have been unable to think of any experimental operations for investigating an *S*'s understanding of the asymmetry of a relation, which differ from operations used to train him to discriminate the relation.

[36] Connectedness is a property of the number series in which Russell was principally interested. It is not, however, a property of all series: for example, it is a property of positions on a scale (no two positions being identical), but it is not necessarily a property of objects measured against a scale (since two of these may be the same). In the experiment reported, the *S*s were presumably comparing lengths of objects, not positions on the length scale.

A particular kind of series is specified by its generating relation. Thus, the relation "heavier than" gives rise to a series of weights. The relations "after" or "before" give rise to temporal series. The relation of class inclusion gives rise to a class hierarchy (e.g., Greeks: men: mortal beings: etc.). Similarly, in a position sequence, such as that involved in the order-discrimination task, the relevant relation is spatial: "left of," or "right of." The asymmetric transitive relation involved in the length-measurement task is also a spatial relation, but a different one, namely, "longer than" (or "shorter than"). In order, therefore, to predict that the development of a grasp of the transitivity of the relation "longer than" will be associated with the genesis of discrimination of position sequences ordered with respect to the relation "left of" or "right of," it is first necessary to show that the relation "longer than" can be defined in terms of the relations "left of" and "right of." However, the relation "longer than" can indeed be so defined: the statement "OA is longer than OB" is a verbal shorthand for some more awkward statement such as, "When OA and OB are juxtaposed so that their left hand extremities, O, are coincident, then the relative positions of their other extremities will be such that OA overlaps OB, i.e., such that A is to the right of B." Thus, relative length is defined in terms of overlapping positions, and it is therefore possible to treat the relation "longer than" as a special case of the left-right relation.

It can therefore be predicted that discrimination of position order and the inferential response $A > B . B > C . \supset . A > C$ will be found to develop in association in children, or—to say the same thing in different words—the age at which the relation "longer than" becomes transitive for children should be closely related to the age at which different serial arrangements of a set of objects are distinguished.

A machine model. The derivation of this prediction might be represented more concretely in the following manner. Suppose we construct a machine which will perform the inference $A > B . B > C . \supset . A > C$. Into one end of this machine we feed the relations between A and B, and between B and C; the machine performs the inference and we read off its deductions concerning the relations between A and C. Now let us construct a second machine which will perform the order discriminations of the order task. Russell's logical analysis of order informs us that there is no operation performed by the first machine which the second machine cannot also perform. Moreover, the first machine can perform certain operations of the second machine: it can discriminate the sequence of three objects.

In the two previous sections, some similarities between the behavior of children and of these hypothetical machines were investigated as a function of age. Logical analysis demonstrates that these two machines are in important respects two versions of the same machine. Consequently, when an *S* behaves like one of these machines, he should also behave like the other. One may therefore make the two following predictions: First, the ages of first appearance of order discrimination and of the inferential response $A > B . B > C . \supset . A > C$ will be found to be closely related. Since the age of first appearance is difficult to establish more than approximately, the second prediction is more significant: there should be an extremely high correlation between the two tasks.

Empirical Evidence

Figure 7 portrays the increase with age of the percentage of *S*s who made an inferential response either in Phase 2 or Phase 3 (i.e., without or with the supplementary verbal instructions) of the measurement task. Juxtaposed with this is the graph showing the increase with age of the percentage of *S*s who reached the criterion in any of the series of the order task. The latter curve is identical to that for Series C with three colors per card, since all *S*s

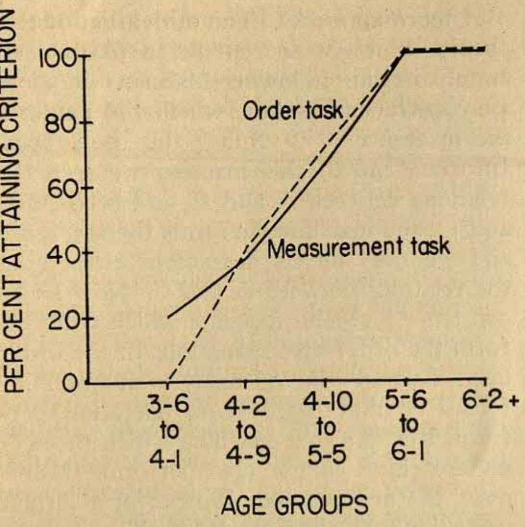


FIG. 7. Percentage of *S*s making an inferential response in the measurement task and discriminating order. The number of *S*s in each age group was 5, 13, 9, 8, and 5, respectively, for the measurement task, and 5, 12, 11, 11, and 5, respectively, for the order-discrimination task.

who reached the criterion in any series were successful in Series "C."

It can be seen that the course of development is almost exactly the same on the two tasks. The youngest age group is the only one at which there is a difference of more than a few percentage points between the two curves and this difference is caused entirely by one *S* who made an inferential response in the measurement task but failed to show any concept of order.[37]

The fact that the two developmental curves shown in Fig. 7 are almost superimposable confirms the prediction, but it would be dangerous to attach too much significance to it. Future investigation may well demonstrate that with other experimental procedures, either or both of the responses studied can be elicited in children of average intellectual capacity younger than these *S*s. If such should be the case, the close correspondence between the two curves would be revealed as fortuitous.

More significant is the fact that there was a strong tendency for *S*s either to fail both tasks or to pass both. This tendency occurred in all age groups. Altogether 37 *S*s took both tasks. Of these, 12 failed both, 22 showed some measure of success in both, 2 made an inferential response in the measurement task but completely failed to discriminate order, and one failed to make an inferential response but reached the criterion in some of the series of the order task. Even when one eliminates the two oldest age groups, in which 100% of *S*s passed both tasks, and the youngest age group, in which almost all *S*s failed both tasks, the correlation remains high. Considering, then, only the children aged 4–2 to 5–5, it was found that of the 20 children who took both tasks, 9 failed both, 9 were successful in both, 1 made an inferential response but showed no concept of order, and 1 showed some concept of order but failed to make an inferential response. Chi square = 12.8 ($p < .0001$), and the correlation (tetrachoric) is .96 ($\phi = .80$). In fact, in view of the impossibility of constructing tests of perfect reliability, the correlation appears to be as high as can be expected in careful psychological research.

Thus both predictions are substantially fulfilled. Discrimination of position order and the inferential response $A > B . B > C . \supset . A > C$ were found to have the same course of development with age, at least in this sample of children and under the experimental conditions employed. Also, children who made an inferential response in the measurement task were also found to discriminate order, and vice versa.

---

[37] The age of this child was 3–8 and her adjustment to the order task was so unusual that the task cannot be regarded as a fair test of order discrimination for her. After the measurement task she discovered that other *S*s, who had not eaten the candy they had won, had had their earnings wrapped up for them in a piece of paper. Every piece of candy she won she then wanted wrapped up at once in a similar piece of paper and placed some distance away from her so that she would not be tempted to eat it immediately. Her attention was focused much more on what was going to happen to the candy than on the task. Once she had won three or four pieces responding randomly she was content and wanted to leave the conflictful experimental situation. Little significance can be attributed to her failure to reveal any concept of order.

Discussion

Colleagues with whom these results have been discussed have sometimes made the point that some degree of correlation between the tasks would be expected, since success on both tasks is probably correlated with intelligence test scores. In purely statistical terms, this point is no doubt valid. However, these tasks were designed in order to test a prediction derived from Piaget's theory of the development of intelligence and, in terms of Piaget's theory, success on them, by definition, reflects intelligence. It therefore begs the question to invoke intelligence to explain the correlation found, since the nature of intelligence is itself the subject of the investigation.

For this reason, intelligence tests were not administered to the *S*s. One would, of course, expect that success on both tasks would correlate with mental age. The kind of inference demanded in the measurement task has been regarded since Aristotle as a manifestation of intelligence; stringing a chain of beads in a given order is a recurrent test item on the Binet.

It is known that correlations between individual test items at the preschool level are typically low relative to correlations between items at higher age levels (Goodenough, 1954, p. 466). Considering the age of the *S*s, the correlation found is thus higher than one would expect on the general grounds that all tests of intellectual functioning correlate. However, the size of the correlation obtained is difficult to evaluate statistically because the sample is small and because it is not possible to establish confidence limits for a tetrachoric correlation or phi-coefficient.

One may note that Piaget's theory does not predict that all tests of intellectual functioning, paired indiscriminately, would correlate with each other. Between success in two carefully designed tasks such as the ones used here, one of which demands a grasp of the transitivity of the spatial relation "longer than," and the other of which demands the discrimination of sequences ordered with respect to an associated spatial relation, it predicts a very high correlation. But there appears to be nothing in Piaget's theory which would predict a particularly high correlation between success in these two spatial tasks and, for example, success in any task used to explore the development of operations of classification.[38, 39]

Conclusion

In the correlation found between order discrimination and the inferential response $A > B . B > C . \supset . A > C$, a prediction derived from a postulate implicit in Piaget's theory has been empirically verified. Within the restricted context of this study, the data confirm the validity of the postulate that when the definitions of logical operations are interdepedent, the isomorphic intellectual processes develop in association. Whether this postulate can be verified over a much wider range of children's intellectual responses and thus provide a solid empirical basis for Piaget's theory of the development of intelligence is a matter for further research.

AGES OF TRANSITION IN THE DEVELOPMENT OF INTELLIGENCE?

So far Piaget's theory has been examined only within the relatively limited context of the experimental work reported in the previous chapters.

---

[38] It is interesting to speculate that the distinction between the operations of the class and spatial interpretations of the logical calculus in Piaget's theory might provide a theoretical basis for the verbal and spatial factors usually found by factor analyses. It still appears to be a matter of interpretation in factor analysis whether the so-called verbal factor is truly a linguistic factor, or whether it relates to reasoning about class membership, relations between classes, etc., which provide the content of so many verbal tests.

[39] An unpublished study of the development of the class-inclusion relation, employed many of the *S*s used in the experiments reported here. There was no obvious correlation between solution of the "class-inclusion" problem and success in the length-measurement and order-discrimination tasks. Unfortunately, the small number of *S*s and the variety of factors found to be related to solution of the class-inclusion problem greatly limit the significance of this finding.

Piaget's theory affirms that in the course of development there are ages of transition during which intelligence changes from relatively "primitive" to more "mature" forms. Such important changes in the nature of intellect should be manifest in longitudinal studies of the development of intelligence as measured by intelligence tests. In the past twenty years there have been several studies of this kind, of which probably the best known is the Berkeley growth study (Jones & Bayley, 1941). In conclusion, the results from these studies are summarized and Piaget's theory is related to the evidence from them.

In the past there have been two principal viewpoints about the development of intelligence. The classical view is that there is no change in the nature of intelligence during development, that there are no qualitative differences in thought and reasoning between early childhood and later years, and that intellect develops through steady and continuous quantitative increments from birth to, at least, adolescence. This viewpoint found its clearest expression in the doctrine of the "constancy of the IQ." At the time of the heyday of the classical view about 25–30 yr. ago, the exponents of the opposite viewpoint were largely limited to Piaget's group at Geneva, although the psychoanalytic view has this in common with Piaget's: the theoretical dichotomy between the "primary processes" characteristic of infancy and the "secondary processes" which are thought to develop approximately at the time of resolution of the Oedipus complex necessarily implies a transition in the development of intellect.

Many of the early criticisms of Piaget's work were advanced before the classical viewpoint had been generally discredited. These criticisms, therefore, tended to assume that there were no changes in the nature of intelligence during development and consequently where Piaget saw a research problem, his critics saw either no problem or a pseudo-problem. Thus, not only were Piaget's early theories (Piaget: 1926, 1928, 1930a, 1930b) severely criticized, but also, unfortunately, the problems on which he was working were largely disregarded. Consequently, most of these critical studies (Bruce, 1941; Deutsche, 1937; Huang, 1943; Huang & Lee, 1945; Isaacs, 1930; Mead, 1932; Oakes, 1946)[40] have tended to be unfruitful; they have led neither to alternative solutions nor to reformulations of Piaget's research questions.[41] They have largely confined themselves to demonstrating, often convincingly, that Piaget's theories, as he stated them, had serious flaws and that the evidence adduced by him was often a good deal less than cogent. Piaget has himself admitted the validity of some of the criticisms of his early work (Piaget, 1953b, pp. 12–13).[42]

The following section summarizes the evidence from longitudinal studies for the existence of changes in the nature of intelligence during early childhood, which are not changes in quantity alone.

Longitudinal Studies of Intelligence

During the last twenty years there has been increasing evidence that scores on psychological tests administered during infancy have little or no predictive value for later intelligence. Table 4 summarizes, as a function of age at initial testing and of test-retest interval, the correlations obtained by 14 investigators who have followed groups of preschool children over a period of years into later childhood. These correlations are remarkably consistent. Tests administered prior to 2 yr. of age sometimes show low but significant correlations over short periods, but the relationships appear to vanish when the test-retest interval is increased.

It is not until around age 3 that significant (but usually low) correlations with scores in adolescence are obtained, and a high relationship with adolescent intelligence is not consistently found until the late preschool and first school years.

---

[40] There have also been generally confirmatory studies (Dennis, 1943; Dennis & Russell, 1940; Nagy, 1948; Russell, 1940). A number of studies of egocentric speech are reviewed by McCarthy (1954).

[41] With at least one important exception (Vigotsky, 1939).

[42] However, his admission is ambiguous: it seems to be retracted later in the same paragraph.

TABLE 4

SUMMARY OF TEST-RETEST CORRELATIONS OBTAINED IN LONGITUDINAL STUDIES OF INTELLECTUAL DEVELOPMENT

Investigator	*N*	Tests Used[a]	Test and Retest Age	*r*
Nelson & Richards (1938, 1939)	48	G × MP	6 mo. × 2 yr.	.37*
Bayley (1949, 1955)	27	C.1 × CPS	6 mo. × 3 yr.	.13
Cattell (1940)	49	C × SB	6 mo. × 3 yr.	.34*
Furfey & Muehlenbein (1932)	27	LH × SB	6 mo. × 4⅔ yr.	−.11
Anderson (1939)	91	G et al. × SB	6 mo. × 5 yr.	−.06
Bayley	27	C.1 × WB	6 mo. × 18 yr.	−.24
Wittenborn (1956)	196	G × SB	3–14 mo. × 4–9 yr.	.05
Nelson & Richards	45	G × SB	1 yr. × 3 yr.	.33
Cattell	100	C × SB	1 yr. × 3 yr.	.56*
Bayley	27	C.1 × CPS	13 mo. × 3 yr.	.41
Furfey & Muehlenbein	27	LH × SB	1 yr. × 4⅔ yr.	−.20
Anderson	91	G et al. × SB	1 yr. × 5 yr.	.05
Bayley	27	C.1 × WB	13 mo. × 18 yr.	−.14
Skodak & Skeels[b]	43	KB × SB	13–21 mo. × 5½–8 yr.	.31
Skodak & Skeels[b]	51	KB × SB	13–21 mo. × 11–16 yr.	.18
Honzik (1938); Honzik, MacFarlane, & Allen (1948)	200	CPS × CPS	21 mo. × 5 yr.	.39*
Honzik et al.	150	CPS × WB	21 mo. × 18 yr.	.07
Anderson	91	G et al. × SB	2 yr. × 5 yr.	.45*
Cunningham (1934)	20	KB × SB	2 yr. × 8 yr.	.24
Bayley	27	CPS × WB	2 yr. × 18 yr.	.39
Driscoll (1933)	39	KB × SB	2–3 yr. × 5–9 yr.	.66*
Goodenough & Maurer (1942)	26	MPS × SB	1½–3 yr. × 7½ yr.	.26
Goodenough & Maurer	21	MPS × SB	1½–3 yr. × 12½ yr.	.18
Bradway (1944, 1945)	52	SB × SB	2–3½ yr. × 13 yr.	.66*
Katz (1941)	268	SB × SB	3 yr. × 5 yr.	.62*
Ebert & Simmons (1943)	181	SB × SB	3 yr. × 6 yr.	.55*
Honzik et al.	200	CPS × SB	3 yr. × 6 yr.	.57*
Ebert & Simmons	181	SB × SB	3 yr. × 10 yr.	.50*
Honzik et al.	200	CPS × WB	3 yr. × 18 yr.	.35*
Bayley	27	CPS × WB	3 yr. × 18 yr.	.40
Skodak & Skeels[b]	34	KB × SB	2½–3½ yr. × 11–16 yr.	.51*
Goodenough & Maurer	28	MPS × SB	3–4 yr. × 12½ yr.	.61*
Ebert & Simmons	181	SB × SB	4 yr. × 10 yr.	.59*
Bayley	27	CPS × WB	4 yr. × 18 yr.	.52*
Bradway	86	SB × SB	4–5½ yr. × 15 yr.	.67*
Bayley	27	SB × WB	6 yr. × 18 yr.	.77*
Honzik et al.	200	SB × WB	7 yr. × 18 yr.	.71*

[a] Abbreviations for tests: G—Gesell Developmental Schedule; MP—Merrill-Palmer Performance Scale; C.1—California First-Year Mental Scale; CPS—California Preschool Scale; C—Cattell Infant Intelligence Scale; LH—Linfert-Hierholzer Infant Scale; MPS—Minnesota Preschool Scale; KB—Kuhlmann-Binet Scale; SB—Stanford-Binet Scale (1916 or 1937 revision); WB—Wechsler-Bellevue Intelligence Scale.

[b] Skodak and Skeels do not report correlations by age groups; the writer has computed these correlations from their published data (Skodak & Skeels, 1949, Table 17, pp. 122–24).

* Significant at the .01 level.

Several reasons have been proposed for the lack of correlation of early test scores with those of later ages. Low reliability cannot be a factor since after 6 mo. of age reliability is usually adequate (Cattell, 1940, p. 49). Escalona (1950) has suggested that the predictive power of infant tests could be improved by adjusting scores on the early tests according to clinical judgments, so that the scores represent not simply the sum of the items passed but clinical estimates of the child's "potential" at the time of testing. She presents some evidence that somewhat higher correlations might be obtained by this method over a short period of years. However, Wittenborn (1956) and his collaborators divided their *S*s into two groups. In one of these the infant intelligence classification was "unqualified"; in the other, the examiner expressed lack of confidence in the results. The correlations with later intelligence were about equally low in both groups.

Another explanation that has been advanced for the inconsistency of early test scores is that the infant tests do not contain suitable kinds of items. Catalano and McCarthy (1954) have suggested that measures of the prelinguistic vocalizations of infants might predict later intelligence better than the infant scale items. The early vocalizations of their 23 *S*s were sampled at an average age of 13 mo., and intelligence tests were administered after an average interval of 31 mo. The eight correlations reported varied from .04 (frequency of back vowels) to .45 (number of consonant types). They appear to be of about the same discouragingly small size as the correlations between comparable ages summarized in Table 4.

Several investigators (Bayley, 1955; Bradway, 1945; Maurer, 1946; Nelson & Richards: 1938, 1939) have performed item analyses of the infant scale test items. Bayley and her associates used the adolescent IQs of their 40 *S*s as the criterion variable. They selected the infant test items which most clearly differentiated the six brightest and six dullest children in adolescence. The total scores for the whole group of infants on these items were then correlated with the intelligence test scores of the same children at 14–16 yr. Bayley reports, "We were unable to get significant correlations even though our sample was composed in large part of the cases on whom the items were selected, including all of the extreme cases that would determine a relationship." She concludes, "I am inclined to think that the major reason for this failure rests in the nature of intelligence itself" (Bayley, 1955, p. 807).

This conclusion suggests strongly that different factors are involved in tests at different ages in early childhood and that there may be ages of transition during which the factors involved in later intelligence develop in children. Hofstaetter (1954) performed a factor analysis of the interage correlations of the Berkeley growth study and found the variance could be accounted for by three factors, with negligible residuals (see Table 5). To the first factor, Hofstaetter gives the name "sensori-motor alertness." He suggests the name "persistence" for the second factor, and relates it to the influence on the test results of such characteristics as cooperation and negativism. The third factor, Hofstaetter says "represents what is commonly called intelligence." Since these factors are derived from correlations between total scores on omnibus intelligence tests, no factors comparable to the spatial, verbal, etc., factors found in more conventional factor analyses can be expected.

Some caution must be exercised in interpreting these factors in terms of psychological functions. To equate one of them with an "ability" involves the dangerous assumption that there are no individual differences in the rate of increase of the ability over the period of high factor loading: otherwise, the test placement of the children with respect to each other would alter and the factor would not appear. It is not clear whether Hofstaetter conceives of "sensori-motor alertness" as an "ability."

TABLE 5

Factor Loadings (After Rotation) of Intelligence Tests at Various Ages, Reproduced From Hofstaetter's[a] Analysis of the Interage Correlations of the Berkeley Growth Study

Factor	Test Age in Months												
	2	5	8	11	14	21	32	48	72	108	144	180	210
No. 1	45	77	88	87	80	53	35	03	00	02	08	00	08
No. 2	−48	−28	−13	15	46	54	65	57	36	07	02	00	05
No. 3	03	−04	06	18	10	39	50	68	90	97	96	98	94

[a] Hofstaetter, 1954, Table 2.

At this stage of knowledge, it seems wiser to refrain from interpreting these factors in terms of specific functions. They should probably be regarded in a more general manner as indicating that the kinds of processes which are involved in passing test items vary from age to age in a more or less regular manner during the preschool period. The precise nature of these processes can only be established by further research, which will not necessarily employ factor-analytic methods.

The third factor does not exert a preponderant influence until 4 yr. of age and the loading of this factor increases considerably between 4 and 6 yr. However, the *S*s of the Berkeley study constitute an intellectually superior group. From 6 yr. of age onwards, successive measures of their mean IQ on different tests range from about 116 to 132.[43] One would expect the intellectual development of a group of children of more nearly normal intelligence at maturity to be slower; moreover, the third factor may be related to the kind of test item rather than to age, and an average child is examined on a given item at a later age than one who is superior. It is therefore quite possible that in a normal group the high loadings on the third factor might emerge at a somewhat later age than in the *S*s of the Berkeley study.[44]

The evidence suggests that there are probably several periods of transition during which the nature of intellectual functions change. Whatever function or group of functions is measured by the infant tests becomes less and less important as age increases. Whether these functions actually decline in importance, or whether the content of the later tests is simply irrelevant to them, is an open question. If there is no decline of these functions then, at least theoretically, it is possible to construct tests at later ages which would correlate with the infancy tests.[45]

Piaget's Levels and the Evidence From Tests

When one attempts to assess Piaget's views in the light of the above material, one finds some things that are predictable from his theory and others that are not. Among the latter one notes that there appear to be periods of transition which he does not mention.

In two works, Piaget (1952, 1954) has discussed in considerable detail the cognitive development that takes place during the first two years of life, his data consisting of observations made on his own three children. His conception of the child's world at birth seems in many ways similar to James' famous concept of a "booming buzzing confusion." The main achievement of the sensori-motor development of the first two years appears to be the transformation of this confusion into a relatively stable world of concrete objects in which the child situates himself. Piaget apparently believes the child's perceptual organization of his immediate environment to be largely complete before 2 yr. of age. Temporally, this period corresponds fairly well with the period of influence of Hofstaetter's first factor.

There is then an hiatus in Piaget's studies of child development. In his works on infancy he leaves the child 18 mo. to 2 yr. old; in his studies of "operational" thinking the youngest *S*s are about 4 yr. old. The intermediate period of development has not—at least not yet—been extensively studied by Piaget. This hiatus corresponds fairly well to the period of maximum influence of Hofstaetter's second factor.

---

[43] Estimated from Bayley (1955, Fig. 8, p. 812).

[44] As yet, no such intellectually normal group has been followed over a period of years. In all 14 investigations whose results are summarized in Table 4, the mean IQ of the *Ss* was above average, and usually substantially above average, although in most samples the dispersion of the IQs around the group mean was normal. The socioeconomic status and education of the adults with whom the children lived were also above average.

[45] It would be interesting to know if, having been unable to construct infant tests which will predict later intelligence, some of the investigators cited have tried to find out if there are any kinds of items in the later tests which are consistently associated with early precocity. It is conceivable that by this means some light might be thrown on the nature and fate of the processes involved in the infant tests.

In Bayley's sample of children the third factor first becomes noticeable at 21 mo., becomes the most important factor at 4 yr., and achieves a fairly stable ceiling at 6 yr. The long period of emergence of this factor is thus completed by approximately the age at which Piaget claims that "operational" thinking emerges. Even making allowance for the superior intelligence of the Berkeley children, it is obvious that this factor plays a considerable role in intelligence tests at an earlier age than the 6–8 yr. at which Piaget places the intellectual changes he describes. Following Piaget's theory, one would predict that a factor analysis of longitudinal data would yield some factor similar to Hofstaetter's third factor. However, one would not expect this factor to become noticeable so early nor to rise quite so fast. Thus, if Piaget's view of the nature of intelligence is valid, the age levels he reports are probably placed too late. This conclusion is consistent with the findings of the experiments reported. The inferential and order-discrimination responses studied in the earlier sections developed in the sample of children employed (whose average IQ is probably 20–30 points lower than that of the Berkeley group) at approximately the age at which one might expect whatever is measured by Hofstaetter's third factor to exert an important influence on their intellectual performance.

This correspondence suggests that when some of the intellectual processes upon which Piaget's theory bears are studied by more carefully designed methods, the age of transition found is quite consistent with published facts about the development of intelligence.

The longitudinal studies provide no evidence to confirm the existence of the period of transition which Piaget places at about 11–12 yr. of age. No fourth factor appears in Hofstaetter's analysis, which indicates that the rank order of children on omnibus tests does not tend to alter at this age. However, this fact cannot be regarded as definitive proof that there is no qualitative change in intellectual performance at this time, since there may be qualitative changes which are not associated with changes in rank order.[46] Whether or not there are changes in the factorial structure of intelligence at this age manifested in more conventional factor analyses, is another question. One may note that there is little evidence for this in the literature: the evidence is only that the spatial, verbal, and numerical factors become more distinct with increasing age (Garrett, 1946). However, the test batteries used do not usually employ tests properly suited for testing Piaget's view about the transition alleged to take place at age 11 yr.; Piaget believes that it is not until this age that children begin to reason about propositions. It might be interesting to examine as a function of age the factorial structure of a battery which includes tests demanding inferences from propositions.[47] In some tests of this kind which have been administered, children under 11 yr. have shown little success (Moore, 1929).

It is clear from the longitudinal studies that Piaget's view that there is an alteration in the nature of intellectual processes toward the end of the preschool years is not as far-fetched as some of his early critics seem to have believed, although it is also clear that Piaget places too late the age at which the transition takes place. This change can be readily explained in terms of the concept of "levels" of intellectual development and is rather difficult to account for by any other theory. In fact, the longitudinal studies provide probably the strongest available evidence in favor of the type of approach which postulates and attempts to identify, "emergent" intellectual processes during early childhood.

The results from the longitudinal studies do not, of course, prove the validity of Piaget's view of the nature of the transition during the middle and late preschool period. But they do demonstrate that the study of the nature of intellectual processes at different ages is concerned with a real, not a pseudo-problem. The main criticism that can be leveled in retrospect against Piaget's early critics is that they contributed little to this problem.[48] It may be suggested that Piaget's theory of intelligence warrants much more detailed investigation than it has hitherto received.

---

[46] It has, of course, been assumed that the converse is true: when there is a marked change in children's test placement relative to each other, it is reasonable to assume that some important qualitative alteration in performance has occurred.

[47] Moore and Anderson (1954) present an interesting use of the "calculus of propositions" in the design of difficult tests involving inferences from propositions.

SUMMARY AND CONCLUSIONS

Piaget's theory of the development of intellect has been examined in a number of ways. The first and primary purpose was to select two areas of intellectual functioning discussed by Piaget and to subject to detailed investigation his description of the development that takes place during childhood within these areas. This purpose was accomplished in the second and third sections. In the second section, the development in children of certain logical inferences, whose validity is tacitly assumed when length is measured, was studied. In the third section, the development of children's concepts of position order was investigated. In each case Piaget's theory, his research methods, and the nature of his evidence were reviewed and discussed, and an experiment was conducted to evaluate certain weaknesses in his approach, suggested by the discussion.

The second purpose was to discuss Piaget's theory of the development of intelligence, to derive an hypothesis from it and to test this hypothesis. The hypothesis concerned the expected correlation between a child's performance of one of the inferences basic in the logic of measurement, and his possession of a concept of position order, reflected in the discrimination of sequences of objects arranged in different orders.

The third purpose was to assess in a general manner the concordance of Piaget's concept of levels of development with the available data on the development of intelligence as measured by intelligence tests.

The evidence suggests the following conclusions.

1. The inference $A > B . B > C . \supset . A > C$ (where "$>$" is interpreted "longer than"), which has the status of an axiom in the logic of length measurement, is generally available to children at least two years before the age at which Piaget locates its development. Factors other than correct reasoning according to the axiomatic standards of the logic of measurement therefore appear to be involved in his studies. The difference between Piaget's experimental procedures and those used here suggests that these factors are the development of skill and interest in the techniques of length measurement, and perhaps vocabulary development (especially, comprehension of the meanings of words and phrases such as "measure," "same length," etc.). The effect of these factors is probably to conceal the reasoning ability of many of Piaget's *S*s.

2. Children show some concept of position order, in general, at least two years before the age at which Piaget places the development of order discrimination. It was found that the number of items in the sequences to be discriminated is an important variable disregarded by Piaget. It is likely, though not proven, that discrimination of order is facilitated for the younger *S*s and the shorter sequences (i.e., less than seven items) when the task is presented using relatively nonverbal methods.

3. There is nothing in the data which suggests that Piaget's concept of "emergent levels" of development should be rejected. The objection that the apparent "age of emergence" is a function of the particular

---

[48] Of course, Piaget's early work was on egocentricity, animism, etc., and not explicitly on intelligence. However, Piaget always explained these phenomena as manifestations of the nature of thought and reasoning at the age at which he found them. Obviously the development of a more mature form of intelligence (which, as defined by Piaget, seems to be closely related to what the psychoanalysts call "secondary processes," Goldstein calls "abstract attitude," and Lashley, "generalizations of the second order") is likely to be associated with many changes in modes of communication, in the perception of causality, etc. This must be expected whether or not Piaget's specific views on egocentric speech, animism, etc., are valid.

task manipulanda and discriminanda was considered in some detail. This objection is not fatal for a concept of levels although it suffices to show that the problem of establishing levels is somewhat more complicated than Piaget appears to have envisaged. The reason why Piaget's results differ from those obtained here is not, in the writer's opinion, because his notion of levels is inapplicable, but because in his research design he fails to eliminate important variables which are not involved in the definition of the processes he sets out to investigate, and which hinder many of his *S*s from responding to the appropriate cues.

4. Piaget's theory of the development of intelligence predicts accurately the high correlation in the individual child between the availability of the inference $A > B . B > C . \supset . A > C$ and the possession of a concept of position order.

5. The available data on the development of intelligence as measured by intelligence tests suggest that there are periods of transition during the preschool years during which the nature of intellectual functions changes. These data support the notion of phases or levels in the development of intellect and seem rather difficult to account for on any other basis. However, the period of transition most relevant to this examination of Piaget's theory appears to occur during the middle and late preschool years, rather than at approximately age 7 yr., as Piaget reports. The evidence thus suggests, once again, that Piaget probably places too late the development of the intellectual processes upon which his theory bears.

While discussing Piaget's theory of intelligence, it was noted that, although his theory is in some ways admirably precise and clear, his conception of psychological process is at variance with that of modern behavioral theory. For Piaget, intellectual processes are not hypothetical constructs, which require specification in terms of the experimental procedures by which they may be elicited, but realities which need only to be diagnosed through intuitive analysis of an *S*'s behavior and language. In his research his major effort is therefore expended on analysis of his *S*s' methods of solution of problems, rather than on experimental design. This method of approach perhaps leads him to overlook relevant variables, with the result that he is misled as to the age of development of the processes he is investigating. In this respect, one contribution of the present study is that it begins to provide more carefully designed experimental procedures for eliciting a few of the processes upon which his theory bears.

In view of the many criticisms that have been made of Piaget, it should be emphasized that the data reported provide considerable support for his notion of emergent levels and also, within the limited scope of these experiments, for his conception of the emergent processes as "operational" thinking.

REFERENCES

Anderson, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Develpm.*, 1939, **10**, 203–212.

Andrew, G., & Harlow, H. F. Performance of Macaque monkeys on a test of generalized triangularity. *Comp. Psychol. Monogr.*, 1942, **19**, No. 3.

Arieti, S. Special logic of schizophrenic and other types of autistic thought. *Psychiatry*, 1948, **11**, 325–338.

Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.

Bayley, Nancy. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, **75**, 165–196.

Bayley, Nancy. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818.

Bergmann, G., & Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.

Binder, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.

Bingham, H. C. A definition of form. *J. anim. Behav.*, 1914, **4**, 136–141.

Bingham, H. C. Visual perception of the chick. *Behav. Monogr.*, 1922, **4**, No. 4.

Boole, G. *An investigation of the laws of thought.* (1854 ed.) New York: Dover, 1953.

Bradway, K. P. IQ constancy in the revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217.

Bradway, K. P. Predictive value of Stanford-Binet preschool items. *J. educ. Psychol.*, 1945, **36**, 1–16.

Bruce, M. Animism vs. evolution of the concept "alive." *J. Psychol.*, 1941, **12**, 81–90.

Campbell, N. R. *Physics: The elements.* Cambridge: Cambridge Univer. Press, 1920.

Catalano, F., & McCarthy, Dorothea. Infant speech as a possible predictor of later intelligence. *J. Psychol.*, 1954, **38**, 203–209.

Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1940.

Cunningham, B. V. Infant IQ ratings evaluated after an interval of seven years. *J. exp. Educ.*, 1934, **3**, 84–87.

Dennis, W. Animism and related tendencies in Hopi children. *J. abnorm. soc. Psychol.*, 1943, **38**, 21–36.

Dennis, W., & Russell, R. W. Piaget's questions applied to Zuni children. *Child Develpm.*, 1940, **11**, 181–187.

Deutsche, J. M. The development of children's concepts of causal relations. Univer. Minnesota, Inst. Child Welf. Monogr. Series. Minneapolis: Univer. Minnesota Press, 1937.

Driscoll, G. P. The developmental status of the preschool child as a prognosis of future development. *Child Develpm. Monogr.*, 1933, No. 13.

Ebert, E., & Simmons, K. The Brush Foundation study of child growth and development. I. Psychometric tests. *Monogr. Soc. Res. Child Develpm.*, 1943, **8**, No. 2.

Escalona, Sibylle. The use of infant tests for predictive purposes. *Bull. Menninger Clin.*, 1950, **14**, 117–128.

Evarts, E. V., & Nissen, H. W. Test of "the abstract attitude" in chimpanzees following ablation of prefrontal cortex. *AMA Arch. Neurol. Psychiat.*, 1953, **69**, 323–331.

Fields, P. E. Studies in concept formation. I. Development of the concept of triangularity by the rat. *Comp. Psychol. Monogr.*, 1932, **9**, No. 2.

Fields, P. E. Studies in concept formation. IV. A comparison of white rats and raccoons with respect to their visual discrimination of certain geometric figures. *J. comp. Psychol.*, 1936, **21**, 341–355.

Fraisse, P., & Vautrey, P. La perception de l'espace, de la vitesse, et du temps chez l'enfant de cinq ans. I. L'espace et la vitesse. *Enfance*, 1952, **5**, 1–20. (a)

Fraisse, P., & Vautrey, P. La perception de l'espace, de la vitesse, et du temps chez l'enfant de cinq ans. II. Le Temps. *Enfance*, 1952, **5**, 102–119. (b)

Furfey, P. H., & Muehlenbein, J. The validity of infant intelligence tests. *J. genet. Psychol.*, 1932, **40**, 219–223.

Garrett, H. E. A developmental theory of intelligence. *Amer. Psychologist*, 1946, **1**, 372–378.

Gellerman, L. W. Form discrimination in chimpanzees and two-year-old children. I. Form (triangularity) per se. *J. genet. Psychol.*, 1933, **42**, 3–27.

Goldstein, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.

Goldstein, K., & Scheerer, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).

Goodenough, Florence L. The measurement of mental growth in childhood. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd ed.) New York: Wiley, 1954.

Goodenough, Florence L., & Maurer, K. M. *The mental growth of children from two to fourteen years: A study of the predictive value of the Minnesota Preschool Scales.* Minneapolis: Univer. Minnesota Press, 1942.

Harlow, H. F. Primate learning. In C. P. Stone (Ed.), *Comparative psychology.* (3d ed.) New York: Prentice-Hall, 1951.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

HOFSTAETTER, P. R. The changing composition of "intelligence": A study in T-technique. *J. genet. Psychol.*, 1954, **85**, 159–164.

HONZIK, M. P. The constancy of mental test performance during the preschool period. *J. genet. Psychol.*, 1938, **52**, 285–302.

HONZIK, M. P., MACFARLANE, JEAN W., & ALLEN, L. The stability of mental test performance between two and eighteen years. *J. exp. Educ.*, 1948, **17**, 309–324.

HUANG, I. Children's conception of physical causality: A critical summary. *J. genet. Psychol.*, 1943, **63**, 71–121.

HUANG, I., & LEE, H. W. Experimental analysis of child animism. *J. genet. Psychol.*, 1945, **66**, 69–74.

HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).

HUNTER, W. S. The question of form perception. *J. anim. Behav.*, 1913, **3**, 329–333.

ISAACS, N. Review of J. Piaget, *Traité de logique. Brit. J. Psychol.*, 1951, **42**, 185–188.

ISAACS, SUSAN. *Intellectual growth in young children.* London: Routledge & Kegan Paul, 1930.

JONES, H. E., & BAYLEY, NANCY. The Berkeley growth study. *Child Develpm.*, 1941, **12**, 167–173.

KATZ, E. The constancy of the Stanford-Binet IQ from three to five years. *J. Psychol.*, 1941, **12**, 159–181.

KREEZER, G., & DALLENBACH, K. M. Learning the relation of opposition. *Amer. J. Psychol.*, 1929, **41**, 432–441.

LANGER, SUSANNE K. *An introduction to symbolic logic.* (2nd ed.) New York: Dover, 1953.

LASHLEY, K. S. The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.

LASHLEY, K. S. Cerebral organization and behavior. Address to the Association for Research in Nervous and Mental Diseases, New York, 1956.

MCCARTHY, DOROTHEA. Language development in children. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd ed.) New York: Wiley, 1954.

MAURER, K. M. *Intellectual status at maturity as a criterion for selecting items in preschool tests.* Minneapolis: Univer. Minnesota Press, 1946.

MEAD, MARGARET. An investigation of the thought of primitive children with special reference to animism. *J. Royal Anthrop. Inst.*, 1932, **62**, 173–190.

MOORE, O. K., & ANDERSON, S. B. Modern logic and tasks for experiments on problem solving behavior. *J. Psychol.*, 1954, **38**, 151–160.

MOORE, T. V. The reasoning ability of children in the first years of school life. *Stud. Psychol. Psychiat., Catholic Univer. of Amer.*, 1929, **2**, No. 2.

NAGEL, E. Measurement. *Erkenntnis,* 1930, **2**, 313–333.

NAGY, M. The child's theories concerning death. *J. genet. Psychol.*, 1948, **73**, 3-27.

NELSON, V. L., & RICHARDS, T. W. Studies in mental development: I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years. *J. genet. Psychol.*, 1938, **52**, 303–325.

NELSON, V. L., & RICHARDS, T. W. Studies in mental development: III. Performance of twelve-months-old children on the Gesell schedule, and its predictive value for mental status at two and three years. *J. genet. Psychol.*, 1939, **54**, 181–191.

OAKES, M. F. Children's explanations of natural phenomenon. *Teachers' Coll. Contr. Educ.*, 1946, No. 926.

PIAGET, J. *The language and thought of the child.* New York: Harcourt, Brace, 1926.

PIAGET, J. *Judgment and reasoning in the child.* New York: Harcourt, Brace, 1928.

PIAGET, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930. (a)

PIAGET, J. *The child's conception of the world.* New York: Harcourt, Brace, 1930 (b)

PIAGET, J. *Les notions de mouvement et de vitesse chez l'enfant.* Paris: Presses Universitaires, 1946.

PIAGET, J. *Traité de logique.* Paris: Colin, 1949.

PIAGET, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1950.

PIAGET, J. *The origins of intelligence in children.* New York: Int. Universities Press, 1952.

PIAGET, J. How children form mathematical concepts. *Scientific Amer.*, 1953, **189** (5), 74–79. (a)

PIAGET, J. *Logic and psychology.* Manchester: Manchester Univer. Press, 1953. (b)

PIAGET, J. *The construction of reality in the child.* New York: Basic Books, 1954.

PIAGET, J. *De la logique concrète de l'enfant à la logique propositionelle de l'adolescent.* Paris: Presses Universitaires, 1955.

PIAGET, J., & INHELDER, BARBEL. *La génèse de l'idée de hasard chez l'enfant.* Paris: Presses Universitaires, 1953.

PIAGET, J., & INHELDER, BARBEL. *The child's conception of space.* London: Routledge & Kegan Paul, 1956.

PIAGET, J., INHELDER, BARBEL, & SZEMINSKA, ALINA. *La géométrie spontanée de l'enfant.* Paris: Presses Universitaires, 1948.

PIAGET, J., & SZEMINSKA, ALINA. *The child's conception of number.* London: Routledge & Kegan Paul, 1952.

RIOPELLE, A. J., ALPER, R. G., STRONG, P. M., & ADES, H. W. Multiple discrimination and patterned string performance of normal and temporal-lobectomized monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 145–149.

ROBINSON, J. S. The sameness-difference discrimination problem in chimpanzee. *J. comp. physiol. Psychol.*, 1955, **48**, 195–197.

RUSSELL, B. *Introduction to mathematical philosophy.* (2nd ed.) London: Allen & Unwin, 1920.

RUSSELL, R. W. Studies of animism: II. The development of animism. *J. genet. Psychol.,* 1940, **56,** 353–366.

SCHOOLEY, M., & HARTMANN, G. W. The role of insight in the learning of logical relations. *Amer. J. Psychol.,* 1937, **49,** 287–292.

SKODAK, MARIE, & SKEELS, H. M. A final follow-up of one hundred adopted children. *J. genet. Psychol.,* 1949, **75,** 85–125.

SPEARMAN, C. *The nature of intelligence.* London: Macmillan, 1923.

TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

VIGOTSKY, L. S. Thought and speech. *Psychiatry,* 1939, **2,** 29–54.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasinin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944.

WITTENBORN, J. R. A study of adoptive children: I. Interviews as a source of scores for children and their homes. II. The predictive validity of the Yale Developmental Examination of infant behavior. III. Relationships between some aspects of development and some aspects of environment for adoptive children. *Psychol. Monogr.,* 1956, **70,** Nos. 1–3 (Whole Nos. 408–410).

(Accepted for publication May 29, 1958)

Vol. 73, No. 6 Whole No. 476, 1959

Psychological Monographs: General and Applied

CONGRUENCIES AMONG INTERPRETATIONS OF PSYCHOLOGICAL TEST AND ANAMNESTIC DATA[1]

KENNETH B. LITTLE[2]
Stanford University

AND

EDWIN S. SHNEIDMAN
VA Neuropsychiatric Hospital, Los Angeles

IN A CLINICAL SITUATION, when one uses diagnostic tools such as interview techniques or psychological tests the ultimate goal is usually to present a description of an individual in such a fashion as to facilitate accurate predictions of the tested *S*'s future behavior. This description may be limited to a single statement (such as a diagnostic label) or may be in the form of an elaborate and detailed evaluation of the *S*'s psychological status, including the genesis of his major personality characteristics and some specific recommendations for the modification of those attributes or for the development of new ones. In either case, the procedure is characterized by the intervention of an interpreter (the psychiatrist, psychologist, psychiatric social worker, or whoever) between the behavior of the *S* and the descriptive statements. Although, as Meehl (1954) has indicated, it may be possible to avoid the interpreting middle man, in current clinical practice this is rarely done. Instead, the final product of a diagnostic interview or psychological test examination consists of a series of statements based upon quasi-empirical norms, the personal experiences of the interpreter, deductions from a hypothesized personality structure, and just plain guesses. Consequently the effective validity and the effective reliability of the diagnostic tools are functions of both the instruments and of the users.

The primary purpose of the present study was to investigate the congruencies among personality descriptions made by clinicians when such descriptions are based on different sources of information (instruments). In particular, the study was designed to investigate congruencies (*a*) for several kinds of *S*s—psychotics, neurotics, psychosomatic reactions, and psychiatrically normal; (*b*) for several *instruments*—the Rorschach technique, the Thematic Apperception Test, the Make A Picture Story test, the Minnesota Multiphasic Personality Inventory, and a psychiatric anamnesis; and (*c*) for several kinds of *interpretive tasks*—true–false factual items, true–false inferential items, *Q*-sort items, diagnostic labels, and ratings of adjustment.

Previous investigations of the congruencies[3] of the instruments considered here have usually been of two major kinds. In the first, every attempt was made to minimize the influence of the interpreter through the use of objective scoring systems (even though the clinicians occasionally crept in through the back door by means of rating scales). These studies have the advantage of presenting precise quantitative results amenable to statistical manipulations, but they have been subject to the criticism that they distort the way in which the test instruments are actually used in the clinical situation. The second type of study has treated the interpreter as a relevant part of the situation and has made comparisons of the similarities and the differences of molar descriptions based on different diagnostic techniques. These latter investigations are unable to measure with any assurance the degree of correspondence of one interpretation with another, based as they are on global matching procedures. The present study attempts to combine the advantages and to circumvent some of the shortcomings of both types of study mentioned above in an investigation of the effects of the psychological test used, the clinician who interprets the test results, and the *S* who is tested.[4]

---

[1] This investigation was supported by a research grant from the National Institute of Mental Health, Public Health Service, under the sponsorship of the University of Southern California.

[2] Now at the National Institutes of Health, Bethesda, Md.

[3] At this point, some concern with the definition of such concepts as "reliability" and "validity" might have been appropriate, but the authors have advertently chosen to place the discussion of these terms after the description of the procedure and the presentation of the results.

SUBJECTS AND PROCEDURE

Test Subjects

Twelve test subjects (Test *S*s) were used, all hospitalized at the time of testing. Three of the *S*s were categorized as "normal" or nonpsychiatric (Test *S* 1, 2, and 3), three as neurotic (Test *S* 4, 5, and 6), three as psychophysiological disorders (Test *S* 7, 8, and 9), and three as psychotic (Test *S* 10, 11, and 12). Selection of the *S*s was based upon an initial screening of records of consecutive new admissions into two Veterans Administration hospitals: a general medical hospital and a neuropsychiatric hospital. Formal requirements in the initial screening were that all *S*s be Caucasian, male, Protestants, native-born veterans of World War II, with an education no less than the 10th or more than the 12th grade, and between 20 and 35 years old. In addition, for the neurotics and psychotics it was required that they had not received shock treatment, that the entrance diagnosis be unequivocally psychotic or neurotic, and that they be in sufficient contact for psychological testing. For the nonpsychiatric (normal) *S*s it was required that they have no record of any previous psychiatric disorder and that the hospital diagnosis be a straightforward medical one with no functional aspects. For the *S*s in the psychophysiological category, in addition to the lack of recorded previous emotional disorder, it was required that the hospital diagnosis be that of a psychophysiological disorder generally considered to be of functional origin.

After the initial screening, the prospective Test *S*s were administered the Cornell Selectee Index, certain subtests from the Form I Wechsler-Bellevue Intelligence Scale (Digit Span, Block Design, and Similarities), and were interviewed by a clinical psychologist. All *S*s with prorated IQ's less than 90 or greater than 115 were eliminated. Prospective nonpsychiatric *S*s who answered items on the Selectee Index which indicated previous emotional disorders or who were, in the judgment of the psychologist, emotionally disturbed were also eliminated. All remaining *S*s were then interviewed by a psychiatrist. In the case of the nonpsychiatric *S*s, the psychiatrist's judgment had to concur with that of the psychologist as to lack of any emotional disturbance before the *S* was retained; if not, the *S* was dropped. For the neurotics, the decision was required that the disorder be definitely nonpsychotic and reasonably representative of one of the classical neurotic categories. For the psychotics, the decision was required that the disorder be definitely nonneurotic and representative of a functional psychosis.

All Test *S*s were administered the Rorschach (Ror), the Thematic Apperception Test (TAT), the Make A Picture Story test (MAPS), and the group form of the Minnesota Multiphasic Personality Inventory (MMPI). All the protocols (other than the MMPI) were recorded on tape and transcribed verbatim. All the testing was done by one clinical psychologist, who had had seven years of clinical and testing experience.

In addition to the test materials, a comprehensive psychiatric case history was obtained from each of the 12 *S*s. All the histories were obtained by one psychiatrist. Four to eight interviews of 1–3 hr. duration were held with the *S*s to obtain the case histories. These histories ranged from 16 to 60 double-spaced typewritten pages and included such materials as presenting problem, developmental history, physical and laboratory findings, description of the family life, etc.[5]

The age, marital status, occupation, hospital diagnosis, etc. for each of the twelve Test *S*s are given in Table 1. An abbreviated case summary for each *S* is presented in Appendix A.

---

[4] There is also a fourth possible source of variance: the different test administrators. In the present study, this aspect was held constant in that all the tests administered to all the *S*s were given by one test administrator. Also, it can be mentioned at this juncture that all the case histories were obtained by one interviewer. This opportunity is taken to extend our appreciation to Phillip A. Goodwin, and Seymour Pollack, who administered all the psychological tests and obtained all the case history data, respectively, for this study.

[5] Limitations of space prevent the duplication of the full case histories here, but case history for Test *S* 10 is presented as "A Case History of Paranoid Schizophrenia" in Burton and Harris (1956).

TABLE 1

TEST SUBJECTS[a]

S	Age	Extraction	Marital Status	Occupation	Hospital Diagnosis
1	25	English-Swedish	Div.	Postal clerk, salesman, policeman	Fractures, left elbow, left tibia, left fibula. No psychiatric disorder.
2	23	Irish	Sgl.	Student at agriculture school	Fractures, right elbow, right hip, right ankle, pneumothorax. No psychiatric disorder.
3	23	Dutch	Sgl.	IBM repairman	Herniorrhaphy. No psychiatric disorder.
4	27	English-Dutch	Mar.	Store salesman	Chronic anxiety reaction, with hysterical conversion neurosis and depression.
5	29	Scotch-Dutch	Mar.	Mechanic	Anxiety reaction, a motor tension state, with symptomatic alcoholism.
6	33	Swedish	Mar.	Store salesman	Mixed psychoneurosis with obsessive-compulsive trends, with phobic reactions.
7	31	English-Dutch	Mar.	Waiter, aspiring composer	Gastro-intestinal irritable colon syndrome.
8	29	English-Irish	Sgl.	Carpenter	Psychophysiologic gastro-intestinal reaction with peptic ulcer syndrome.
9	26	French-English	Sgl.	Machinist	Psychophysiologic skin reaction manifested by atopic dermatitis.
10	22	Dutch-English	Sgl.	Student	Schizophrenic reaction, paranoid type.
11	29	English	Sgl.	Unskilled labor	Schizophrenic reaction, chronic, undifferentiated type.
12	26	Danish	Sgl.	Unemployed	Schizophrenic reaction, primarily catatonic in type.

[a] All *S*s are male, Caucasian, Protestant, native born, 10th–12th grade education, 90–115 IQ, and between the ages of 22 and 33.

Test Judges

Forty-eight clinical psychologists served as test judges (TJ) for the study. Twelve TJ's were selected to interpret the Rorschach (R1–R12), 12 to interpret the TAT (T1–T12), 12 to interpret the MAPS test (M1–M12), and 12 to interpret the MMPI (I1–I12). Selection of the MMPI judges was from lists of clinical psychologists provided by Robert E. Harris and Starke R. Hathaway; MAPS test judges were selected from a list provided by Edwin S. Shneidman; Rorschach and TAT judges were selected from lists drawn up by the authors in consultation with the editors of the *Journal of Projective Techniques.* The forty-eight TJ's constituted, therefore, a sample of skilled interpreters, rather than a random sample of clinical psychologists in general. The names of the 48 TJ's are listed, in alphabetical order, in Appendix B.

Anamnesis Judges

Twenty-three psychiatrists and one psychologist served as evaluators of the anamnestic materials. Eleven were practicing psychoanalysts, members of the American Psychoanalytic Association, one was a

Diplomate in Clinical Psychology (A1–A12), and 12 were, by their own evaluation, nonanalytically oriented members of the American Psychiatric Association (P1–P12). Selections were from lists prepared in consultation with the officers of the local associations. The 24 anamnestic judges (AJ) are listed, alphabetically, in Appendix B.

In the selection of both TJ's and AJ's an initial number somewhat larger than the final group was contacted and asked to participate. The study was described briefly and it was indicated that an honorarium of fifty dollars would be paid for participation. The final 72 judges were selected at random from those accepting.

Design and Procedure

Each TJ evaluated the test protocols of four Test *S*s, one from each of the four subject categories. The only information provided the judges in addition to the protocols was that the *S*s were Caucasian, male, veteran, Protestant, with high school education, and between 22 and 33 years of age.[6] The order in which any four protocols were evaluated was a random one with the restriction that each protocol be once the first one evaluated, once the second, etc. Since there were 12 judges for each test type, four judges evaluated *S*s T1, 4, 7, 10 (Set I), four evaluated *S*s T2, 5, 8, 11 (Set II), and four T3, 6, 9, 12 (Set III). These same sets were used for all test types.

Ten days after the conclusion of the evaluation of the four protocols, each TJ repeated the analysis of the first record of the four protocols he had seen, and, as another task, made an evaluation of an hypothetical individual (a "normal," a psychophysiological reaction, a neurotic, or a psychotic stereotype) corresponding to one of the four Test *S* categories. (The stereotype evaluations were for another study and are not considered further in this monograph.)

The AJ's evaluated two *S*s on the basis of the anamnestic materials. They also performed a stereotype evaluation but did not, however, do a repeat analysis.

Interpretive Task

In order that comparisons might be made among the various TJ's and AJ's, the evaluation of the Test *S* on the basis of the materials was structured in the form of five tasks for the judges:

1. Answer the question, "How would you label this person diagnostically?"
2. Indicate on a line scale the degree of maladjustment of the *S*.
3. Complete a *Q* sort of 76 items describing behavioral reactions to social stimuli and general adjustive patterns.
4. Answer 117 true and false items representing the type of statements made in psychological test reports and also indicate the level of confidence in each answer.
5. Answer 100 true and false items of a more or less factual nature about the *S*'s past and present life and also indicate the level of confidence in each answer.

The items of Tasks 2, 3, 4, and 5 are presented in Appendix C.

The 76 items of the *Q* sort consisted of 48 items describing positive, negative, or neutral reactions to positive or negative stimuli from either other persons or the environment in general. The remaining 28 items were equally divided among statements of satisfactory general adjustment reactions and unsatisfactory adjustment reactions.

The 117 psychological report type items were selected from a population of 1706 such items, derived primarily by abstracting approximately 80 test reports on patients in neuropsychiatric and general medical hospitals and mental hygiene clinics. Inasmuch as these items were predominantly about negative personality characteristics, an equal group of statements was constructed, asserting a logical contradictory of each of the original items.[7] From this combined group of items a random sample of 500 items was selected and assigned, by one of the authors, into one of 18 psychological test report categories. These 18 categories are presented in Appendix D.

---

[6] It may be appropriate to indicate at this point that every effort was made to give no extraneous clues to the judges; e.g., the authors—who were then in a VA neuropsychiatric hospital setting—conducted their correspondence with the TJ's and AJ's on university stationery and with a university mailing address.

[7] The two logical contradictories should be mutually exclusive. For example, a logical contradictory of the statement "He hates his mother" is the statement "He has no strong feelings toward his mother."

A recategorization of the 500 items by the other author resulted in disagreement on only 16 items, and these 16 were then eliminated. From the remaining 484 items the final sample of 117 was selected by proportional sampling among the categories, following the percentages of items in each category indicated in a previous study.[8]

The 100 true–false, more or less factual items were a sample from an initial list of 223 statements which could be asserted to be factually either true or false about each *S* on the basis of the anamnestic and hospital record data—e.g., "He was an only child"—and which were presumably relevant to personality description or development.

RESULTS

The results of this study will be discussed under the following six general headings: Diagnostic Statements, Degree of Maladjustment, *Q* Sorts, True–False Personality Items, True–False Factual Items, and Reliability. In each of these sections the results will be described and the conclusions from, and implications of, these results will be discussed.

Diagnostic Statements

The first task of each TJ and each AJ was to indicate the diagnostic label he would assign to the Test *S* whose material he was evaluating. These diagnostic statements made by the judges are presented, verbatim, in Table 2. Although a complete statistical interpretation of this table cannot be given, some comments about its qualitative features can be made. The outstanding impression (a disappointing one) is that diagnostic labels given by the TJ's from the psychological tests were remarkably varied and inaccurate—if one takes seriously psychiatric diagnosis and the authors' categorization of the patient. The second outstanding feature is that this inaccuracy is particularly noticeable in the case of the nonpsychiatric *S*s.

These data have some interesting implications and lead to several speculations. One such speculation is that the training of clinical people tends to emphasize the pathological. (Only recently has there been any real attention to assets and to "ego strengths.") It is an unhappy fact that very few clinicians ever work with "normal" subjects and that clinicians as a group have no population of normals to use as an internalized reference from which to draw inferences. It is also relevant to indicate that inaccuracy of diagnostic labeling does not necessarily imply inaccuracy in describing psychodynamics, personality traits, etc. A psychological defense may have positive as well as negative value; a coping mechanism is not necessarily a deformity. There is an important difference between stating, on the one hand, "An individual has obsessive features" and, on the other hand, "An individual is an obsessive neurotic." The TJ's apparently made errors of confusing these two.

Certainly the data indicated in Table 2 give everyone interested in the use of psychological techniques some pause for thoughtful reflection.[9]

---

[8] These 18 categories, and the proportion of items in each category, came from Shneidman, Walther, and Little (1951), in which the authors developed this category set based on 21 test reports which had been analyzed into 1210 discrete items of personality description. It should be mentioned at this point that the present study (involving 12 *S*s of four different categories) is a direct outgrowth of this former study which involved only one *S*, and thus raised many questions which could be approached only by multiplying the types of *S*s and by having more than one *S* in each type. Two studies, one on the validity of MMPI interpretations (Little & Shneidman, 1956) and one on the validity of thematic interpretations (Little & Shneidman, 1955) based on the one case (and evaluated by several judges) have already been published.

[9] The findings presented in Table 2 raise the question—as in the case of the Holtzman and Sells (1954) study where projective techniques were not particularly successful in making predictions of flying success—that it may well be that the psychological tests were used either inappropriately or in a way in which clinicians ordinarily do not use tests. It is felt that to ask clinicians to make diagnoses, ratings of adjustment, to do *Q* sorts, and to complete T–F questionnaires are not especially inappropriate tasks for clinicians, but the present study does raise the entire issue of the role of the "blind diagnoses." This is cited, not to give any opinion on one side or another, but simply to indicate that the topic is most relevant to an understanding of the present data.

TABLE 2

DIAGNOSTIC LABELS ASSIGNED TO 12 TEST *S*s BY 48 TJ'S AND 24 AJ'S

	Judge
Test *S* 1 (Nonpsychiatric)	
ROR	1. Passive dependent personality 2. Anxiety neurosis 3. Hysterical personality; passive aggressive 4. Schizophrenic with homosexual tendencies
TAT	1. Psychopath—immaturity reaction 2. Schizophrenic; hebephrenic and paranoid features 3. Aggressive reaction 4. Simple schizophrenia or alcoholic
MAPS	1. Schizoid character 2. Schizoid personality; organic brain damage? 3. Schizophrenic, chronic 4. Psychopathic personality with some schizoid trends
MMPI	1. Character neurosis 2. Chronic sub-optimum adjustment level 3. Hysteria 4. An extremely defensive normal
Anamnesis	1. No psychotic disease or disorder 2. Neurotic personality, mild 3. Character disorder, mild 4. Phallic-narcissistic character disorder
Test *S* 2 (Nonpsychiatric)	
ROR	1. Schizoid character—depressive trends 2. Compulsive character 3. Passive aggressive personality disorder 4. Conversion reaction
TAT	1. Psychoneurosis, mixed, homosexuality 2. Anxiety neurosis 3. Neurotic, obsessive-compulsive 4. Character neurosis
MAPS	1. Passive aggressive (dependent) personality 2. Neurotic reactive depression 3. Neurotic (hysteric) 4. Anxiety reaction with compulsive features
Test *S* 2 (Nonpsychiatric)—Continued	
MMPI	1. Normal 2. Hysterical character 3. Without personality disorder 4. Psychoneurosis, conversion hysteria reaction
Anamnesis	1. Within normal range—no psychiatric disorder 2. No psychiatric disease 3. No psychiatric disorder present 4. No psychiatric illness
Test *S* 3 (Nonpsychiatric)	
ROR	1. Incipient schizophrenic 2. Schizophrenic, probably paranoid type 3. Paranoid schizophrenia 4. Schizoid personality, paranoid and homosexual trends
TAT	1. Schizophrenic 2. Anxiety neurosis with hysterical features 3. Character neurosis 4. Passive-aggressive reaction with schizophrenic episodes—emotional instability
MAPS	1. "Normal" with mild obsessive and compulsive trends 2. Schizophrenic, mixed type, chronic, severe 3. Schizophrenic reaction, catatonic features 4. Character disorder with hysterical and homosexual trends
MMPI	1. Passive-aggressive personality (asocial psychopath) 2. Hysterical personality 3. Simple adult maladjustment 4. Simple adult maladjustment, emotional instability
Anamnesis	1. Character neurosis, passive-aggressive type 2. Character within normal limits 3. Normal 4. Normal

TABLE 2—Continued

Test *S* 4 (Neurotic)	
ROR	1. Mixed neurosis with obsessive features 2. Obsessive compulsive, possibly with schizoid trends 3. Character disorder with anxiety hysterical features 4. Neurotic—obsessive compulsive, anxiety and depression conspicuous
TAT	1. Normal adult with character defenses of an obsessive-compulsive nature 2. Character disorder, passive dependent features 3. Passive aggressive reaction with hysterical features 4. Psychotic features in a passive-feminine character
MAPS	1. Immaturity reaction 2. Obsessive compulsive homosexual 3. Obsessive compulsive 4. Hysteric phobic reaction with obsessive trends
MMPI	1. Normal 2. Essentially normal individual, mild obsessional reactions plus homoeroticism problem 3. Normal 4. Mild obsessive with some somatization
Anamnesis	1. Dissociative reaction (hysterical) 2. Anxiety reaction with depersonalization episodes 3. Psychoneurosis, severe, hysterical type 4. Personality trait disturbance with neurotic reaction

Test *S* 5 (Neurotic)	
ROR	1. Paranoid schizophrenic in remission 2. Mixed pre-genital character, passive-aggressive personality, passive-aggressive type 3. Obsessive compulsive reaction 4. Schizophrenic reaction, paranoid type
TAT	1. Schizophrenia, paranoid trends 2. Paranoid schizophrenia in remission or partial remission, with possible acting out aggressive psychopath character structure 3. Schizophrenic reaction 4. Obsessive neurotic with depression
MAPS	1. Schizo-affective psychosis 2. Latent (incipient) schizophrenic reaction 3. Paranoid schizophrenic 4. Schizoid personality with alcoholism (possibly psychotic episodes)
MMPI	1. Simulated or acute psychosis in a basically psychopathic personality 2. Psychopathic personality 3. Manic depressive reaction, manic type 4. Psychosis, manic with paranoid features
Anamnesis	1. Passive-aggressive personality, aggressive type 2. Anxiety state, with alcoholism and paranoid trends 3. Latent schizophrenia 4. Anxiety reaction

Test *S* 6 (Neurotic)	
ROR	1. Anxiety state, moderately severe 2. Anxiety and depression in a compulsive personality 3. Anxiety state (hysterical?) 4. Neurotic personality, hysterical defense mechanisms
TAT	1. Depressive or schizo-affective 2. Obsessive-compulsive neurosis, paranoid trends 3. Normal 4. Passive-dependency reaction with reactive alcoholism—psychoneurosis
MAPS	1. Anxiety reaction, mild, with obsessive compulsive features 2. Hysterical character 3. Could be a normal 4. Normal adult with mild homosexual conflict
MMPI	1. Anxiety neurosis superimposed on a latent paranoid schizophrenia 2. Mild neurotic depression 3. Psychoneurosis, mixed with depression, anxiety, and somatic complaints 4. Anxiety neurosis

TABLE 2—Continued

Test *S* 6 (Neurotic)—Continued	
Anamnesis	1. Psychoneurosis, obsessional, in a schizoid character 2. Severe obsessive compulsive neurosis with depressive trends 3. Mild, chronic, progressive schizophrenia, simple type 4. Mixed psychoneurosis, or schizophrenic reaction, acute, early type, undifferentiated

Test *S* 7 (Psychophysiological)	
ROR	1. A somewhat shy and inhibited normal 2. Obsessive compulsive character 3. Oral character disorder with ego restriction 4. Schizoid, inadequate personality
TAT	1. Immaturity reaction 2. Character disorder, passive-dependent, possibly alcoholic 3. Adjusted normal 4. Neurotic depression, anxiety state in a passive inhibited compulsive character
MAPS	1. Obsessive-compulsive 2. Paranoid schizophrenia 3. Anxiety reaction, possibly latent schizophrenia 4. Schizophrenic reaction, catatonic type
MMPI	1. Psychoneurosis mixed—somatic complaints 2. Psychoneurotic reaction in an immature and self-centered person 3. Hypochondriasis 4. Somatization reaction, probably stomach ulcer
Anamnesis	1. Somatization reaction, character problem 2. Anxiety reaction with gastro-intestinal somatization 3. Passive-aggressive personality with somatization of anxiety 4. Psychogenic gastro-intestinal conversion syndrome (nonpsychotic)

Test *S* 8 (Psychophysiological)	
ROR	1. Immature character, inadequate 2. Masochistic character 3. Personality trait disturbance, passive aggressive personality 4. Anxiety reaction
TAT	1. Schizophrenia mixed or catatonia. Possibly organic factors 2. Anxiety reaction in an inadequate personality 3. Passive dependency reaction 4. Neurosis with marked overt anxiety: phobias? obsessiveness? possible epileptic
MAPS	1. Chronic brain syndrome with psychosis 2. Psychogenic gastro-intestinal reaction 3. Borderline schizophrenia 4. Anxiety reaction with somatic complaints
MMPI	1. Depressive reaction in a chronically anxious and tense person 2. Depressive reaction 3. Reactive or neurotic depression 4. Anxiety neurosis
Anamnesis	1. Psychophysiologic gastro-intestinal reaction with peptic ulcer 2. Psychophysiologic gastro-intestinal reaction with peptic ulcer 3. Psychosomatic gastro-intestinal disturbance, ulcer 4. Anxiety, gastro-intestinal reaction

Test *S* 9 (Psychophysiological)	
ROR	1. Psychoneurosis, mixed type—has some compulsive and hysterical trends 2. Mixed neurosis, hysterical-phobic and anxiety features predominating 3. Psychoneurosis, mixed, with anxiety and depression 4. Anxiety hysteria
TAT	1. Compulsive neurosis, or psychopathic 2. Character neurosis 3. Anxiety reaction, precariously maintained 4. Character disorder with narcissistic and psychopathic features
MAPS	1. "Normal" but with some obsessive features 2. Schizophrenia, mild, chronic, ambulatory 3. Essentially normal with some immaturity that may be largely a function of age 4. Obsessive character disorder, with much tension and anxiety

TABLE 2—Continued

Test *S* 9 (Psychophysiological)—Continued	
MMPI	1. Passive-dependent personality 2. Hysterical personality 3. Adult situational reaction 4. Normal personality
Anamnesis	1. Passive dependent with somatization (skin) 2. Passive dependent character disorder with severe inhibitions 3. Psychophysiologic skin reaction, passive dependent character 4. Passive aggressive personality

Test *S* 10 (Psychotic)	
ROR	1. Ambulatory schizophrenic 2. Compulsive character 3. Obsessive-compulsive personality breaking down 4. Inadequate character
TAT	1. Impulse neurotic, immaturity reaction 2. Character disorder, depressive features 3. Immaturity reaction and behavior disorder, "punishment seeker" 4. Neurosis: some depressive features and probably considerable anxiety
MAPS	1. Anxiety neurosis 2. Neurotic depression 3. Character disorder, psychopathic personality 4. Depressive reaction
MMPI	1. Paranoid schizophrenic 2. Acute agitated psychotic or severely disturbed neurotic 3. Paranoid schizophrenic episode with acute anxiety 4. Paranoid schizophrenic episode with marked anxiety
Anamnesis	1. Schizophrenic reaction, paranoid type 2. Schizophrenic, mixed type, paranoid defenses prominent 3. Schizophrenic, paranoid 4. Paranoid schizophrenia

Test *S* 11 (Psychotic)	
ROR	1. Simple schizophrenic 2. Schizophrenic reaction in partial remission, chronic 3. Psychotic disorder, schizophrenic reaction 4. Schizophrenic reaction, paranoid type
TAT	1. Conversion hysteria 2. Schizophrenic in partial remission, paranoid features 3. Immaturity reaction; impulse neurosis 4. Schizophrenia
MAPS	1. Schizophrenic reaction 2. Normal 3. Simple schizophrenic or schizophrenic character disorder 4. Schizophrenic chronic with paranoid trends
MMPI	1. Psychosis in a schizoid personality 2. Schizo-affective 3. Acute anxiety state 4. Manic depressive psychosis with paranoid features
Anamnesis	1. Schizophrenic episodes in a poorly organized personality 2. Schizophrenic reaction, chronic, severe, undifferentiated 3. Schizophrenic reaction, type undifferentiated 4. Schizophrenic reaction, undifferentiated

Test *S* 12 (Psychotic)	
ROR	1. Passive aggressive character, passive dependent type 2. Neurasthenic with paranoid trends 3. Organic or simple schizophrenic 4. Schizophrenia, paranoid features
TAT	1. Psychosis, probably schizophrenic type 2. Anxiety neurosis 3. Highly schizoid but still obsessive personality 4. Schizophrenic process with decompensating obsessive compulsive defenses
MAPS	1. Schizoid personality 2. Schizoid character with moderate depressive and obsessive features 3. Schizoid personality 4. Passive dependent character disorder with borderline psychotic trends

TABLE 2—Continued

Test *S* 12 (Psychotic)—Continued	
MMPI	1. Reactive depression in a latent paranoid schizophrenic 2. Passive dependent personality 3. Psychoneurosis, obsessive compulsive type with depressive features 4. Paranoid schizophrenic, probably in partial remission

Test *S* 12 (Psychotic)—Continued	
Anamnesis	1. Catatonic schizophrenic 2. Schizophrenic reaction, catatonic type 3. Schizophrenic reaction, catatonic type 4. Schizophrenic reaction, catatonic type

Although a complete statistical evaluation of the material in Table 2 was not undertaken, an estimate of the consistencies of the diagnoses for any given Test *S*, where diagnoses were based upon the same protocol or anamnestic material, was obtained in the following fashion:

1. Each diagnosis for each *S* based upon the same information source (i.e. the Rorschach, the TAT, etc.) was paired with every other diagnosis based upon that same source of information, e.g., the diagnosis of Judge R1 for Test *S* 1 was paired with each of the other Rorschach judges' diagnostic statements for *S*1. Six paired comparisons were thus obtained. A similar procedure was used for the TAT, MAPS, MMPI, and AJ's. For each *S*, then, 30 pairs of diagnostic statements were available.

2. As a chance probability baseline, 23 sets of four diagnoses were selected at random from the 240 listed in Table 2. For each set, all possible pairings of the four diagnoses were made giving 138 chance combinations in all.

3. The total of 498 pairs of diagnoses were typed and distributed to four competent psychodiagnosticians with instructions to rate each pair on a six-point scale of similarity. The instructions to the raters and the defining points of the scale are given in Appendix E. Since six possible scores (0–5) were available for each pair of diagnoses, product moment correlations were computed among all raters. The average correlation among the four psychodiagnosticians was .73 indicating a moderate degree of agreement among them. The chief source of disagreement appeared to be over the use of similarity of dynamics versus similarity of symptoms.

4. For each pair of diagnoses, the sum of the four ratings assigned by the judges was used as the index of similarity of the pair. These ranged from 0 to 20 where 0 indicated unanimous agreement on the part of the raters that the two diagnoses were utterly different and 20 indicated unanimous agreement that they were identical. The six figures thus obtained for the paired combinations of diagnoses for any one *S* based upon a single information source were in turn summed to give an index of similarity for each Test-*S*, information-source cell.

Table 3 gives the mean similarity indices (transformed back to a 0–5 scale) for TJ's

TABLE 3

MEAN DIAGNOSIS SIMILARITY INDICES AMONG FOUR JUDGES GROUPED ACCORDING TO INFORMATION SOURCE AND *S* TYPE

*S*s		Ror.	TAT	MAPS	MMPI	Anam.
Nonpsychiatric	1	1.5	1.7	2.1	2.4	2.7
	2	1.9	2.6	2.5	2.0	4.9
	3	2.9	1.8	1.2	2.2	3.2
Neurotic	4	2.8	1.8	2.4	3.6	3.0
	5	2.0	2.7	2.6	2.1	2.3
	6	3.3	1.2	2.5	3.1	2.2
Psychophysiological disorder	7	2.0	1.5	1.8	2.7	3.3
	8	2.4	1.4	1.4	3.8	4.5
	9	3.8	2.2	1.7	2.2	3.2
Psychotic	10	2.3	2.7	2.3	3.7	4.6
	11	3.2	1.4	1.6	1.9	3.9
	12	1.8	2.1	3.4	2.0	5.0
Mean		2.50	1.76	2.12	2.64	3.57

and Test *S*s. The fifth column presents the same figures for the AJ's.

The first analysis of these data was a two-way analysis of variance of the first four columns of Table 3 with test types as columns and *S* category as the rows. None of the possible effects were significant, i.e., similarity of diagnosis did not differ as a function of either test type or *S* category.

For the second analysis, simple *t* tests of the differences between the means for the TJ's and that for the AJ's were computed. The mean value for the AJ's was significantly higher (beyond the .05 level) in every case.

As a final analysis, each of the mean values for the four tests and for the AJ's were compared against the mean of the randomly selected sets of four diagnoses ($N = 23$). The mean of this distribution was 1.94 and the s was 1.34. The mean values for the Rorschach, MMPI, and the AJ's exceeded this chance value beyond the .05 level; the TAT and MAPS did not.

The results relative to agreement among diagnoses obtained from the above analyses may be summarized as follows:

1. Agreement among diagnoses where the diagnoses are based upon either the Rorschach, the MMPI, or anamnestic material, tend to be greater than chance.
2. Agreement among diagnoses based upon the MAPS or the TAT do not exceed chance agreement.
3. Agreement among diagnoses based upon anamnestic material is significantly greater than that based upon any other source of information.
4. Agreement among diagnoses based upon any of the sources considered do not appear to be affected by the type of *S* evaluated.

As a testament to the efficacy of psychodiagnostic instruments, these results leave something to be desired. Inspection of Table 2 has suggested that diagnostic labels based upon blind analyses of protocols may be quite wide of the mark and the present analysis indicates that the judges may not be even shooting at the same target.

How can one account for such disappointing results? Within the structure of the investigation there are certain factors that tend to influence the agreement index: extra weight was given to a single deviant diagnosis, agreement among the judges of similarity was by no means perfect, and the pool of diagnoses from which the chance index was derived was of restricted range. The first two factors tend to lower the similarity index and increase its standard error while the latter raises the mean chance similarity index. However, even under these conditions, the finding that AJ's did significantly better than TJ's indicates that a higher degree of agreement is possible. Parenthetically, we might add that this higher degree of agreement of diagnoses based on anamnestic data as compared to test data is undoubtedly related to the derivation of our diagnostic labels themselves. They are based upon "the facts" of a man's life, e.g., a schizophrenic is not defined as a person with pure *C* on the Rorschach or a high *Sc* score on the MMPI, but rather as a man who hallucinates, is delusional, manifests autistic thinking, etc.—the very "facts" which are presented in a well-taken anamnesis.

The major implication of the present results is that the blind diagnosis of psychological test protocols for clinical purposes is of dubious value. It is possible that combinations of instruments may produce better agreement, but when tests are used singly, agreement on diagnoses among test interpreters is not to be expected.

Degree of Maladjustment

Each TJ and each AJ was asked to rate each *S* he evaluated on a maladjustment scale. Nine different scores, ranging from 0 through 8, were possible on the maladjustment scale. Four such scores were obtained for each Test *S*. These scores were assembled into three sets corresponding to the three sets of Test *S*s, i.e., *S*s 1, 4, 7, 10; *S*s 2, 5, 8, 11; and *S*s 3, 6, 9, 12, and investigated by means of a three-way analysis of variance (McNemar, 1955, Model XVII). In each analysis, TJ's formed the rows,

TABLE 4

Analysis of Variance of Maladjustment Ratings for Three Sets of Test *S*s

		Variance		
Source	*df*	Set I	Set II	Set III
TJ's	12	3.0	2.4	4.6
Tests	3	2.7	6.5	1.7
S type	3	14.4*	40.6*	17.8*
Test × *S*	9	8.4*	5.3*	6.0*
Remainder	36	2.6	1.8	2.0
Total	63			

* Significant at .01 level.

type of *S* the columns, and tests the blocks. Results of the analyses are given in Table 4. Table 5 gives the mean maladjustment ratings of four judges for each *S* based upon the several information sources including the anamnestic data.

In each of the three analyses, differences among tests were not significant but differences among *S* types were. In addition, the interaction between test type and *S* type was also significant in each analysis. Differences among *S* types were to be expected but not in exactly the form obtained. In *S* Set I, the psychotic was judged most sick, the nonpsychiatric *S* and the psychophysiologic *S* next, and the neurotic as the least ill. In *S* Set II, the order was psychotic, neurotic, psychophysiologic, and nonpsychiatric; and in *S* Set III, psychotic, nonpsychiatric, neurotic, and psychophysiologic. These results may be compared with those of the AJ's who ranked the *S* types the same in all three sets.

TABLE 5

Average Maladjustment Ratings for Each *S* by TJ's and AJ's

*S*s		Ror.	TAT	MAPS	MMPI	Anam.
Nonpsychiatric	1	4.0	6.5	5.0	3.5	1.8
	2	3.5	2.8	3.5	3.0	1.0
	3	6.0	5.0	5.8	3.5	2.0
Neurotic	4	3.5	3.8	4.2	2.0	6.5
	5	4.0	6.0	7.5	6.8	7.0
	6	4.8	3.8	2.2	6.5	6.8
Psychophysiological disorder	7	4.0	3.0	6.2	5.0	3.5
	8	3.0	5.8	6.8	4.5	4.5
	9	3.8	4.5	3.2	3.5	3.5
Psychotic	10	5.0	4.8	5.0	8.0	7.0
	11	7.0	7.2	5.8	7.5	7.0
	12	6.0	6.8	6.2	5.8	8.2

The observation that information-source, *S*-type interaction was significant in all three analyses indicates, of course, variation in the order of ranking depending upon information source, i.e. profile differences. For example, in *S* Set I, the TAT judges considered the nonpsychiatric *S* most maladjusted, whereas the MAPS judges considered the psychophysiologic *S* most maladjusted and the neurotic *S* as the least. To test whether such patterns were stable ones, a fourth analysis was made using the data in the first four columns of Table 5 with *S* sets as rows, *S* type as columns, and tests as blocks. The results of this analysis are presented in Table 6. As would be expected from the preceding analyses, differences among *S* types were significant (in that the psychotics were consistently considered most ill) and there were no significant differences in mean judgments among test types after the combination of the *S* sets. The interaction, moreover, was also nonsignificant, indicating that the apparent test by *S*-type effects observed previously were a function of the particular four Test *S*s used in each of the three sets.

Finally, a two-way analysis of variance (McNemar, 1955, Case VIII) was made of the data in Table 5 for each of the four Test *S* types separately in order to compare

TABLE 6

Analysis of Variance of the Mean Maladjustment Ratings of Each Test *S* by Four TJ's

Source	Sum of Squares	*df*	Variance
Sets	102.9	8	12.8
Test type	22.0	3	7.3
S type	350.8	3	116.9*
Test × *S*	62.7	9	7.0
Remainder	577.1	24	24.0
Total	1115.5	47	

* Significant at .01 level.

TABLE 7

ANALYSES OF VARIANCE OF MEAN MALADJUSTMENT RATINGS OF FOUR TYPES OF *S*s BY TJ'S AND AJ'S

Source	*df*	Variance			
		Non-psychiatric	Neurotic	Psychophysiological	Psychotic
*S*s	2	438.6*	654.4	186.2	115.2
Information source	4	556.2**	321.7	145.1	162.6
Interaction	8	63.8	248.9	125.4	92.1

* Significant at .05 level.
** Significant at .01 level.

the average ratings of the AJ's with that of the TJ's. Table 7 gives the results of these analyses.

Only in the case of the nonpsychiatric *S* are the mean maladjustment ratings of the TJ's significantly different from those of the AJ's. Judges for all test types tended to rate the nonpsychiatric *S*s as significantly more maladjusted than the AJ's.

The results of ratings of maladjustment may be summarized as follows:

1. Psychotics as a group are rated as most maladjusted on the basis of all of the tests used.
2. No significant differences occur among over-all mean ratings of maladjustment for nonpsychiatric, psychophysiologic, or neurotic *S*s.
3. No individual test elicits a significantly greater or lesser maladjustment index than any other test.
4. For nonpsychotic *S*s, maladjustment ratings based on test protocols are significantly higher than those based on anamnestic material.
5. For other than nonpsychiatric *S*s, maladjustment ratings based upon test protocols are not significantly different from those based on anamnestic material.

As a first observation, it is apparent that the maladjustment rating data reflect the same situation observed in relation to diagnoses. Nonpsychiatric *S*s are tarred with the pathologic brush. One is reminded of Ernest Jones' response to the question as to the characteristics of a normal person—that he didn't know; he had never met one.

The present data do point up a phenomenon not tested in the analysis of the diagnoses; psychotics are distinguished from the other three classes of *S*s. It seems probable, therefore, that although agreement among diagnoses is no greater for psychotics than for any other *S*s, the diagnoses used are of a more pathologic nature.

Q-Sorts

The first question asked in analyzing the *Q*-sort results was similar to that raised in connection with the diagnostic statements, i.e., Do judges working with different test types differ in the amount of agreement among themselves about the test subjects? The form of analysis used was the pseudo three-way analysis of variance with *S* sets forming the rows, tests the columns, and *S* types the blocks. Entries in the cells were the average coefficients for the six correlations among the four judges interpreting any particular protocol. None of the possible effects was significant, i.e., there were no differences among over-all average intercorrelations for the different tests nor for the different *S* types. The interaction effect was also nonsignificant.

Table 8 gives the analysis of variance data and Table 9 the average intercorrelations among the *Q* sorts arranged by tests and by Test *S*s. For comparison, over-all averages for tests, *S* types, and the average intercorrelations among the AJ's are also given.

TABLE 8

ANALYSIS OF VARIANCE DATA OF THE AVERAGE INTERCORRELATIONS AMONG TJS' *Q* SORTS

Source	Sum of Squares	*df*	Variance
S sets	11574.3	8	1446.8
S types	1878.7	3	626.2
Tests	3635.2	3	1211.4
S type × Test	2986.1	9	331.8
Remainder	17411.0	24	725.4
Total	37485.3	47	

TABLE 9

AVERAGE *Q*-SORT CORRELATIONS AMONG FOUR TJ'S USING DIFFERENT INFORMATION SOURCES FOR EACH TEST *S*

Test	*S*s												Average
	1	2	3	4	5	6	7	8	9	10	11	12	
Rorschach	56	44	−13	09	36	−06	45	58	−12	−03	64	61	31
TAT	46	14	−02	17	24	−09	51	41	16	23	19	67	27
MAPS	45	28	14	45	22	38	50	56	05	25	07	71	35
MMPI	28	44	23	59	52	39	04	49	21	70	45	08	39
Average	44	33	06	34	34	16	39	52	08	32	36	55	33
Anamnesis	54	72	54	23	60	59	12	42	66	72	40	44	52

The second question asked was whether the TJ's working with a specific test agreed among themselves to a greater extent than between tests. The following method of analysis was used to secure an answer to this question: For each Test *S* the six intercorrelations among the four TJ's working with a specific test were added to the six intercorrelations among the four TJ's working with another test and the average secured. This average was paired with the mean of the 16 correlations *between* the two sets of four judges and the difference obtained. For each pair of tests, 12 such differences were available, one for each Test *S*. A simple *t* test of the mean difference of these 12 correlated differences was then computed.

Table 10 gives the mean differences obtained by the above method. In all cases there was slightly greater agreement within tests than between, but only in the case of the MMPI paired with each of the other tests were the differences significant, i.e., agreement between MMPI judges and the other TJ's was significantly lower than the agreement among judges working within a test type. Among the projective technique judges, however, agreement within tests was not, on the average, significantly higher than across tests.

The third question raised was the amount of agreement between the TJ's and the AJ's. This is, of course, the question of validity where psychiatric judgments are used as the criterion. To answer the question, the four correlations of each TJ with the four AJ's were transformed into z's and averaged. The average correlations (after back transformation) are given in Table 11. Those significantly different from zero at the .01 level ($\sigma_{\text{av}\ z} = .058$) are marked with an asterisk. Examining these values in conjunction with the average intercorrelation among the AJ's themselves (Table 9), it is apparent that part of the low agreement can be accounted for on the basis of disagreement among criterion judges. The values are, however, rather distressingly low in spite of this mitigating factor.

To test differences among tests in correlation with the AJ's, three pseudo three-way analyses of variance were made, one for

TABLE 10

MEAN DIFFERENCES IN AVERAGE INTERCORRELATIONS BETWEEN TWO TESTS COMBINED AND THEIR CROSS CORRELATIONS FOR 12 *S*s

Test	Ror.	TAT	MAPS
TAT	.012		
MAPS	.027	.037	
MMPI	.087*	.176*	.187*

* Significant at .02 level.

TABLE 11

Average Correlation of Four TJ's With Four AJ's on *Q* Sort

S		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	−10	−38*	−39*	24*
	2	25*	−14	−17*	54*
	3	−05	−07	−11	11
Neurotic	4	13	17*	23*	18*
	5	41*	34*	39*	54*
	6	22*	00	−01	26*
Psycho-physiologic disorder	7	02	09	02	03
	8	47*	38*	46*	41*
	9	08	05	11	09
Psychotic	10	07	−07	21*	38*
	11	32*	24*	30*	32*
	12	20*	53*	53*	16*

* Significant at .01 level.

each *S* set. Table 12 gives the results of these analyses. In each, TJ's formed the rows, *S* types the columns, and test types the blocks. In all, differences among *S* types were significant, but in none were differences among tests significant. For Set I, the interaction between test and *S* type reached the 5% level of significance, but since the effect did not appear in the other two analyses it must be considered an artifact of the particular group of *S*s in that set.

Although differences among *S* types were significant in each *S* set, the specific type with the highest mean correlation with the AJ's was different from set to set. In order to determine over-all effects, a fourth analysis was made using test types as rows, *S* types as columns, and *S* sets as blocks. Entries in the cells were the average correlations of four TJ's with four AJ's for one *S*. Table 13 gives the results of this analysis. As would be expected, the Test *S* × *S* set interaction was significant, i.e., in different *S* sets differing *S* types elicited the highest agreement with the AJ's. Block effects were also significant indicating differences in the over-all agreement between TJ's and AJ's for the different *S* sets. Highest agreement, for example, occurred between TJ's and AJ's for the *S*s of Set II. Finally, over-all column effects were significant. Testing the differences among the means indicated that this effect was determined by the remarkably low agreement between TJ's and AJ's on the nonpsychiatric *S*s. Differences among means for the other three *S* types were not significant.

A fourth question of interest is the amount of stereotypy in the TJ's evaluation of the *S* protocols. Since each TJ made a *Q* sort for four different Test *S*s, stereotypy may be defined as the difference between the average correlation among the judge's *Q* sorts and some index of the "true" similarity among the four *S*s. The difficulty, of course, lies in securing an index of "true" similarity. For our analysis, we selected the mean of patterned samples of correlations between Test *S*s within an *S*

TABLE 12

Analyses of Variance Data for Mean *Q*-Sort Correlations of TJ's With Four AJ's

Source	*df*	Variance		
		Set I	Set II	Set III
Judges	12	1049.8	647.8	540.7
Tests	3	1596.1	1711.0	47.5
S type	3	2728.5**	1949.9*	4298.5*
Test × *S* type	9	841.6*	657.5	47.3
Remainder	36	350.7	453.6	1451.6
Total	63			

* Significant at .05 level.
** Significant at .01 level.

TABLE 13

Analysis of Variance of the Average *Q*-Sort Correlations of Four TJ's with Four AJ's

Source	*df*	Variance
Tests	9	4472.6
S sets	2	51551.3**
S type	3	23597.1**
S set × *S* type	6	6150.0*
Remainder	27	2061.9
Total	47	

* Significant at .05 level.
** Significant at .01 level.

set but across judges within one test type. For example, on Set I, all possible correlations among the *Q* sorts of TJ's R-1 for TS-1, R-2 for TS-4, R-3 for TS-7, and R-4 for TS-10 were computed. To these were added the intercorrelations among the *Q* sorts of R-1 for TS-4, R-2 for TS-7, R-3 for TS-10, and R-4 for TS-1; and R-1 for TS-10, R-2 for TS-1, R-3 for TS-4, and R-4 for TS-7. The average of these 24 correlations formed the baseline estimate of the "true" similarity among the four Test *S*s for the Rorschach judges of Set I. Similar indices were computed for other TJ's and for other *S* sets. Since this baseline estimate might still include a specific bias due to test type, an initial two-way analysis of variance was made of the baseline average correlations using *S* sets as rows and tests as columns. There were no significant differences among tests or among *S* sets.

A generalized bias resulting from a tendency of all clinical psychologists to say similar things about different *S*s may still exist in the baseline estimates. However, no method of analysis eliminating this bias was available.

TABLE 14

Average Correlations Among Four *Q* Sorts for Four Different *S*s by Each TJ

	Ror.	TAT	MAPS	MMPI
S Set I	−04	−06	22	−07
	26	−05	36	03
	41	−10	18	07
	28	22	26	−01
Mean baseline	228	000	255	005
	211	*−036*	*228*	*−012*
S Set II	35	38	16	00
	36	10	08	−09
	21	02	24	06
	08	24	18	−02
Mean baseline	250	185	165	−012
	193	*095*	*181*	*−065*
S Set III	25	15	28	56
	31	−10	48	38
	34	−07	10	09
	30	03	04	26
Mean baseline	300	002	225	322
	041	*000*	*135*	*145*

Table 14 presents the average correlations among *Q* sorts for four *S*s for each TJ arranged by *S* set and test type. In the fifth row for each set is given the average of the mean values for each of the four judges in a given *S* set-test cell and in the sixth row, in italicized figures, the mean inter-*S* interjudge correlation used as a baseline. A two-way analysis of variance of the differences between the values in the fifth rows and sixth rows was made. Table 15 presents the results. Test type effects were not significant, i.e., there was no greater tendency for one test to elicit stereotyped *Q* sorts than any other. Subject set effects, however, were significant, suggesting that one group of *S* protocols (Set III) tended to elicit more stereotypic evaluations than the other two. Since TJ's and Test *S*s were assigned at random to the three sets, no immediate explanation for this phenomenon occurs to us.

As a last point, a simple test of the deviation of the mean of the difference correlations from zero for all tests and *S* sets combined was significant, i.e., a statistically significant degree of stereotypy exists among the TJ's. The absolute magnitude, of course, is quite small although examination of Table 14 indicates that certain judges have quite high intercorrelations among their four *Q* sorts.

The results of the analyses of the *Q* sort data may be summarized as follows:

1. Agreement among judges interpreting the same protocol is not significantly affected by the tests.

TABLE 15

Analysis of Variance of the Differences Between Mean Within-Judge *Q*-Sort Correlations and Mean Across-Judge *Q*-Sort Correlations for Four Different *S*s

Source	*df*	Variance
S sets	2	2598.5*
Tests	3	362.3
S set × Test type	6	88.5
Total	11	

* Significant at .01 level.

2. Agreement among judges interpreting the same protocol is not significantly affected by the nosological category of the *S*.

3. Among projective tests, the degree of agreement among judges interpreting the same protocol is not significantly higher than the agreement with judges interpreting a different projective technique protocol for the same *S*.

4. Agreement between interpretations of MMPI protocols and projective technique protocols for the same *S* is significantly lower than agreement among judges interpreting the same protocol.

5. The amount of agreement between test interpretations and judgments of psychiatrists is not significantly different for the four tests used in this study.

6. The amount of agreement between test interpretations for nonpsychiatric *S*s and psychiatric judgments is significantly lower than that for neurotics, psychophysiological disorders, and psychotics.

7. Test interpreters tend to make their interpretations in a stereotyped manner independent of the *S*.

8. Nosological category of the Test *S* does not significantly affect the degree to which stereotyped interpretations of test protocols are made.

The fact that the degree of agreement among the judges interpreting specific projective technique protocols is no higher than their agreement with judges interpreting different projective test protocols for the same *S*s probably implies a common psychodynamic approach to the interpretation of projective test protocols. This common approach seems to pervade interpretation to such an extent that unique contributions of the tests may be masked. The fact that the interpretation of MMPI protocols tend to be more in agreement among themselves than with the interpretations of projective techniques may well be a function of the following characteristics of the MMPI: that the data presented in the MMPI protocols have a fairly high degree of specificity; that there is less variability as to what MMPI (as compared with projective test) TJ's use for cues; that MMPI TJ's have fewer cues than the projective test TJ's; that conclusions drawn from MMPI data are less inferential and more close to the phenomenologically given (in that the MMPI is more similar to the anamnesis); and that, whereas the projective techniques obtain their intellectual permissiveness from psychoanalytic theory, the MMPI, on the other hand, operates within an empirical tradition and deals largely with more denotable aspects of behavior. Little wonder then that the "vision" of the clinicians who deal with questionnaire techniques (like the MMPI) is less blurred than that of the clinicians who must infer from projective techniques. However, it should be pointed out that the amount of agreement between test interpretations, on the one hand, and judgments from the anamneses, on the other hand, was not significantly different for *any* of the four tests used in this study.

It should be stated that the findings and implications of the *Q* sorts (and T–F items) should be clearly differentiated in the reader's mind from the findings and interpretations of the diagnoses and degree of maladjustment. That is to say, the rather disappointing findings that the tests do not do well in obtaining either inter-judge agreement or agreement between test judges and psychiatric judgments in terms of diagnostic labels does not necessarily mean that the tests do not retain a useful and important function in delineating psychodynamics and characteristics of personality as evidenced in *Q*-sort items and T–F items.

True–False Personality Items

Every TJ and AJ had (in addition to the diagnoses, estimates of maladjustment, and *Q* sorts), indicated whether he considered each of 117 psychological report type items true or false for each *S*. The analysis of these data was limited to comparisons between the judgments of the TJ's based upon test protocols and those of the AJ's based upon the anamnestic materials.

The initial step in the analysis was to tabulate the judgments of the four AJ's on the 117 items for each Test *S* and to select those items in which 3 of the 4, or 4 of 4 AJ's agreed. Table 16 gives the number and percentage of the 117 items for each *S* on

TABLE 16

Number and Percentage of 117 Items Marked the Same by 3 Out of 4 or 4 Out of 4 Psychiatrists

S	No.	%
1	87	74
2	111	95
3	94	80
4	85	73
5	101	86
6	99	85
7	84	72
8	99	85
9	105	90
10	105	90
11	102	87
12	99	85

which 3 out of 4 or 4 out of 4 agreements occurred. A modal list was then constructed using the decision of the AJ's for the selected items as the criterion judgment. (Items were not included when the AJ's were evenly split on their truth or falsity.)

Tetrachoric correlations were used as the index of agreement between each TJ and the criterion. Four such indices were available for each TJ, one for each *S* evaluated.

As with the *Q*-sort material, three pseudo three-way analyses of variance were made of the present material, one for each *S* set. In each, TJ's formed the rows, *S* types the columns, and tests the blocks. Table 17 gives the analysis of variance data for the three analyses of variance.

TABLE 17

Analysis of Variance Data for the Tetrachoric Correlations of TJ's With the Modal Judgment of the AJ's on Report Type Items

Source	*df*	Variance		
		Set I	Set II	Set III
TJ's	12	1325.7	875.3	557.8
Tests	3	718.2	1073.3	2025.7*
S type	3	2567.0	10786.6**	16487.4**
Test × *S* type	9	2894.7*	23875.8*	3118.4
Remainder	36	1095.6	1022.3	1607.2
Total	63			

* Significant at .05 level.
** Significant at .01 level.

TABLE 18

Average Tetrachoric Correlations of Four TJ's With the Modal Judgment of Four AJ's on Report Type Items

*S*s		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	−20	04	17	59
	2	−10	−20	12	47
	3	−31	−32	−19	38
Neurotic	4	34	36	39	−02
	5	74	42	64	22
	6	67	44	04	79
Psycho-physiologic disorder	7	38	42	45	14
	8	72	75	65	55
	9	58	18	44	22
Psychotic	10	40	26	28	81
	11	65	37	32	80
	12	62	85	67	44

The results varied from set to set. In Set I only the interaction approached significance, the main effects did not. In Sets II and III there was a significant *S*-type effect, an interaction effect in Set II and a test effect in Set III. It is rather difficult to draw conclusions from such data, so the independent scores from the separate judges were averaged for a composite "validity" index. Table 18 gives these mean correlations with the criterion, of the four TJ's working with each *S*-test combination.

A fourth pseudo three-way analysis of variance was made of the data in Table 18 using Test *S*s as rows, tests as columns, and *S* types as blocks. The results of this analysis, presented in Table 19, indicate significant differences in agreement with the psychiatrists for different *S* types and a significant interaction effect of the Tests × *S* type. The differences among *S* types is largely accounted for by the extremely low amount of agreement between the TJ's impressions of the nonpsychiatric *S*s and the psychiatrists' impressions of these same individuals.

Only one definite conclusion can be drawn from these data: TJ's have a significantly lower correlation with AJ's on nonpsychiatric *S*s than they do on the other three nosological categories. However, inspection

TABLE 19

ANALYSIS OF VARIANCE OF THE AVERAGE TETRACHORIC CORRELATIONS OF TJ'S WITH ANAMNESTIC MODAL JUDGMENT OF REPORT TYPE ITEMS

Source	*df*	Variance
Test *S*s	8	793.0
S types	3	5956.5**
Tests	3	512.9
S × Test type	9	1070.7*
Remainder	24	389.8
Total	47	

* Significant at .05 level.
** Significant at .01 level.

of the average correlations in Table 18 in light of the significant interaction effect reported for Table 19 indicates varying degrees of efficiency of the four tests for different *S* categories. Specifically, the Rorschach seemed about equally efficient for the three *S* categories of neurotic, psychophysiological reaction, and psychotic, and markedly inefficient for the nonpsychiatric *S*s; the TAT showed its optimum efficiency with the psychotic and psychophysiological *S*s; the MAPS test indicated a pattern similar to that of the TAT, and the MMPI gave its best performance for the normal and psychotic *S*s.

It is interesting that in the *Q* sorts there was a significantly higher degree of agreement between TJ's and AJ's for the psychotics than for any other nosological category, whereas with the True–False items there was a significantly lower degree of agreement of the TJ's with the AJ's for the nonpsychiatric *S*s. In other words, there were no differences in the amount of agreement between test results and anamnestic results as far as the *Q*-sort data were concerned among the nonpsychiatric, neurotic, and psychophysiologic disorders, whereas for the True–False data there is no discrimination among the neurotic, psychophysiologic disorder, and psychotic Test *S* data. It would almost appear that the True–False personality items function to discriminate between normal on the one hand and all other ("sick") individuals on the other hand. In relation to the findings for the specific psychological tests, one comment can be made about the MMPI: specifically, that inasmuch as it apparently does differentiate the two extreme groups, its routine inclusion in a battery of tests is probably advisable, not so much to pick out the extremely aberrant individual, but more importantly to focus on the relatively well-adjusted (normal) *S*. Among the projective techniques, the present data would suggest that the Rorschach would be first choice for an additional test in the armamentarium of the clinician—as indeed the surveys of use of tests in clinics demonstrate.

True–False Factual Items

A series of analyses similar to that described for the personality type True–False items was made for the 100 True–False factual items which the TJ's had marked as True or False for each *S*. Criterion keys were established by examining the case histories of each *S* for positive evidence of the truth or falsity of each statement. When no evidence either way was available the item was eliminated. As a result, the total number of items considered varied somewhat from *S* to *S*, ranging from 89 to 100.

Each TJ's marks of true or false for the items from one *S* were matched against the key and the tetrachoric correlation between the two determined. Three pseudo three-way analyses of variance were made of the results, one for each *S* set. Table 20 presents the results of these analyses.

TABLE 20

ANALYSIS OF VARIANCE DATA FOR TETRACHORIC CORRELATIONS OF TJ'S WITH CASE HISTORY KEY FOR FACTUAL ITEMS

		Variances		
Source	*df*	Set I	Set II	Set III
TJ's	12	546.3	265.1	469.1
Tests	3	5815.8**	1255.7*	76.9
S type	3	3089.5**	2781.7**	3387.8**
Test type × *S* type	9	942.0	1150.2**	705.4*
Remainder	36	521.4	235.7	321.1
Total	63			

* Significant at .05 level.
** Significant at .01 level.

In all, differences among *S* types were significant, although the type of *S* most accurately described varied from set to set. In Set I, the psychophysiological *S* was judged most correctly; in Set II, the neurotic; and in Set III, the psychotic. In all three sets the nonpsychiatric *S*s were least correctly described.

In two of the sets (I and II) differences among test types were significant with the MMPI being superior as compared to the projective tests. In Set III, this difference did not occur. Also in two sets (II and III) the interaction of test by *S* type was significant.

A final analysis of the combined sets was also made. Table 21 gives the average correlations that constituted the data for the pseudo three-way analysis. As in the case of the report type True–False items, Test *S*s formed the rows, tests the columns, and *S* types the blocks. Table 22 gives the results of this analysis. None of the possible effects was significant, i.e. there was no difference among mean values for tests, *S* types, or interaction. The apparent superiority of the MMPI disappeared in the combined data.

It is apparent that the results of the analysis of the factual True–False items were inconclusive. One might question at this point whether psychological tests can, or ought, to correlate well with factual case history data. The whole issue of the postdictive function of psychological tests is a debatable one. If one wants to know if an individual was an only child or came from a broken home, etc., a psychological test is not the optimal technique for ascertaining this type of information.

TABLE 21

AVERAGE TETRACHORIC CORRELATIONS OF FOUR TJ'S WITH CASE HISTORY CRITERION FOR FACTUAL ITEMS

*S*s		No. of Items	Ror.	TAT	MAPS	MMPI
Nonpsychiatric	1	96	−12	07	−22	42
	2	99	16	22	27	51
	3	98	−05	02	−20	10
Neurotic	4	98	12	12	09	53
	5	98	42	−10	50	41
	6	97	16	34	37	13
Psychophysiologic disorder	7	95	11	59	45	35
	8	98	−05	10	02	06
	9	99	15	−10	25	06
Psychotic	10	96	14	21	04	56
	11	94	32	04	06	13
	12	88	32	29	29	22

TABLE 22

ANALYSIS OF VARIANCE OF AVERAGE CORRELATIONS OF FOUR TJ'S WITH CASE HISTORY CRITERION

Source	Sum of Squares	*df*	Variance
Test *S*s	5596.0	8	724.5
S type	1709.5	3	569.8
Tests	1788.0	3	596.0
S type × Test type	2999.2	9	333.2
Remainder	6919.3	24	288.3
Total	19012.0	47	

Reliability

The results reported in this section are comparisons of the interpretations made on the first protocol evaluated by the TJ's and interpretations made when the same protocol was re-presented approximately 10 days later (cf. Procedure). The data are in the form of coefficients of agreement between first and second evaluations.

Table 23 gives the two *diagnoses* of each TJ for the Test *S*s classified by test and by *S* type. These 48 pairs of diagnoses were presented to four competent psychodiagnosticians who rated each pair on a 0–5 scale with 5 representing maximum similarity. The average correlation among the four raters for the 48 pairs was .83, indicating a good degree of agreement as to the similarity of the diagnostic pairs. The index of similarity for any specific pair of diagnoses was the mean of the ratings of the four judges. Table 24 gives the indices for each test type. A two-way analysis of variance of the data of Table 24 was made, using *S* type as rows and test type as columns.

TABLE 23

INITIAL AND REPEAT DIAGNOSES OF THE SAME *S* BASED UPON THE PRESENTATION OF THE SAME TEST PROTOCOL

S	Rorschach	TAT	MAPS	MMPI
Nonpsychiatric				
1	(a) Passive dependent personality [initial diagnosis]	(a) Psychopath—immaturity reaction	(a) Schizoid character	(a) Character neurosis
	(b) Inadequate personality, passive—dependency reaction [repeat diagnosis]	(b) Schizophrenic, probably catatonic; psychopath	(b) Schizoid character	(b) Neurotic reaction, very defensive
2	(a) Schizoid character—depressive trends	(a) Psychoneurosis—mixed—homosexuality	(a) Passive aggressive (dependent) personality	(a) Normal
	(b) Schizoid character—depressive trends	(b) Psychoneurosis, mixed, or psychosis, schizophrenia, in remission	(b) Passive—dependent personality	(b) Essentially normal —basically a hysteroid personality
3	(a) Schizoid personality, paranoid and homosexual trends	(a) Passive-aggressive reaction with schizophrenic episodes—emotional instability	(a) Character disorder with hysterical and homosexual trends	(a) Simple adult maladjustment, emotional instability
	(b) Schizoid personality with decompensating features—although some ego is intact	(b) Mixed psychoneurosis, severe, with depressive, phobic, hysterical and passive aggressive features	(b) Passive dependent character neurosis	(b) Emotional instability
Neurotic				
4	(a) Obsessive compulsive possibly with schizoid features	(a) Character disorder—passive dependent features	(a) Obsessive compulsive homosexual	(a) Essentially normal individual, mild obsessional reactions plus homo-eroticism problem
	(b) Obsessive compulsive	(b) Character disorder —hysterical and depressive features	(b) Obsessive compulsive	(b) Anxious, uncertain individual with probably homoerotic problems

TABLE 23—Continued

S	Rorschach	TAT	MAPS	MMPI
	Neurotic—Continued			
5	(a) Mixed pre-genital character, passive-aggressive personality, passive-aggressive type	(a) Paranoid schizophrenia in remission or partial remission, with possible acting out aggressive psychopath character structure	(a) Latent (incipient) schizophrenic reaction	(a) Psychopathic personality
	(b) Anal character with obsessive compulsive and masochistic trends	(b) Depressive reaction in obsessive personality	(b) Schizophrenia (incipient)	(b) Psychopathic personality
6	(a) Anxiety and depression in a compulsive personality	(a) Obsessive-compulsive neurosis, paranoid trends	(a) Hysterical character	(a) Mild neurotic depression
	(b) Severe anxiety neurosis in an immature character. Intellectual and compulsive obsessive defenses too weak to protect him from anxiety	(b) Hysterical character, paranoid trends	(b) Hysterical character	(b) Hysterical personality with mild neurotic depression
	Psychophysiologic Disorder			
7	(a) Oral character disorder with ego restriction	(a) Adjusted normal	(a) Anxiety reaction, possibly latent schizophrenia	(a) Hypochondriasis
	(b) Passive dependency reaction with hysterical features	(b) Fairly well adjusted normal	(b) Schizophrenia	(b) Hypochondriasis (severe psychoneurosis with anxiety and hypochondriacal features predominating)
8	(a) Anxiety reaction	(a) Neurosis with marked overt anxiety: phobias? obsessiveness? possible epileptic	(a) Anxiety reaction with somatic complaints	(a) Anxiety neurosis
	(b) Conversion reaction	(b) Not psychotic. Severe neurosis or character disorder with antisocial behavior	(b) Anxiety reaction with somatic complaints; immaturity personality	(b) Anxiety neurosis

TABLE 23—Continued

S	Rorschach	TAT	MAPS	MMPI
		Psychophysiologic Disorder—Continued		
9	(a) Psychoneurosis, mixed, with anxiety and depression (b) Neurotic or schizoid personality	(a) Anxiety reaction, precariously maintained (b) Character neurosis	(a) Essentially normal with some immaturity that may be largely a function of age (b) Normal, if age is in lower range	(a) Adult situational reaction (b) Psychoneurosis, anxiety reaction, mild
		Psychotic		
10	(a) Inadequate character (b) Obsessive-compulsive neurotic with strong homosexual conflicts	(a) Neurosis; some depressive features and probably considerable anxiety (b) Passive-feminine character disorder with some depressive and some borderline schizophrenic trends	(a) Depressive reaction (b) Phobic reaction with depression	(a) Paranoid schizophrenia episode with marked anxiety (b) Schizophrenia, paranoid type, acute
11	(a) Psychotic disorder; schizophrenic reaction (b) Psychotic disorder, probably a schizophrenic reaction	(a) Immaturity reaction; impulse neurosis (b) Passive dependency reaction (immaturity reaction)	(a) Simple schizophrenia or schizophrenic character disorder (b) Anxiety hysteria	(a) Acute anxiety state (b) Psychotic reaction: schizophrenia
12	(a) Passive aggressive character, passive dependent type (b) Passive-dependent character—possibly alcoholic	(a) Psychosis—probably schizophrenic type (b) Schizophrenia	(a) Schizoid personality (b) Schizoid personality with a strong suspicion of organicity	(a) Reactive depression in a latent paranoid schizophrenic (b) Reactive depression (and/or conversion?) with latent paranoid schizophrenic reaction

None of the possible effects was significant, i.e., there were no significant differences among test types in amount of agreement between repeat diagnoses, nor among *S* types.

Ratings of the *degree of maladjustment* of the *S* were made on second evaluation in the same fashion as on the first. Disagreement between first and second ratings for each judge is indicated in Table 25 in the form of discrepancy scores. A negative number indicates that the judge changed his rating in the direction of increased maladjustment; a positive number indicates the converse, and a zero indicates no change. The changes range from zero to one-half the scale length (8 units). An analysis of variance, however, indicated no significant differences among tests in absolute amount of change of ratings.

TABLE 24

MEAN SIMILARITY INDICES BETWEEN FIRST AND SECOND DIAGNOSES BASED UPON THE SAME TEST PROTOCOLS

Ss		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	4.8	1.8	5.0	3.5
	2	5.0	2.5	5.0	4.5
	3	3.8	1.8	3.5	4.5
Neurotic	4	4.2	3.5	4.0	4.5
	5	2.5	.2	5.0	5.0
	6	3.5	1.8	5.0	3.8
Psycho-physiologic disorder	7	3.0	5.0	2.2	4.5
	8	2.2	2.2	4.2	5.0
	9	2.5	2.2	5.0	2.5
Psychotic	10	1.5	2.0	3.0	4.8
	11	5.0	3.8	.2	1.0
	12	4.5	5.0	3.5	4.2
Mean		3.7	2.6	3.8	4.0

TABLE 25

CHANGES IN MALADJUSTMENT RATINGS FROM FIRST TO SECOND OBSERVATION

Ss		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	0	−1	0	0
	2	−2	−1	0	0
	3	−1	1	0	1
Neurotic	4	1	−3	0	0
	5	0	1	1	−1
	6	0	−2	1	0
Psychophysiological disorder	7	1	0	0	−1
	8	0	−2	−2	1
	9	2	1	0	0
Psychotic	10	0	−2	4	2
	11	1	0	2	0
	12	−4	0	0	1
Mean		1.00	1.17	.83	.58

Table 26 gives the correlations between *Q* sorts of first and second evaluations of the same protocol by each judge and the average for each test. Table 27 gives the analysis of variance data for Table 26. Differences among test types are significant in that the TAT judges have a significantly ($P = .01$) lower mean reliability coefficient than that for judges of the other tests.

Table 28 gives the tetrachoric correlations between first and second evaluations of the

TABLE 26

Q-SORT RELIABILITY FIGURES (PRODUCT MOMENT CORRELATION) FOR THE 48 TJ's

Ss		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	66	61	94	62
	2	76	45	52	73
	3	70	17	19	51
Neurotic	4	67	58	76	59
	5	36	−01	80	88
	6	26	−07	50	79
Psychophysiological disorder	7	70	78	47	58
	8	51	32	69	84
	9	69	17	74	66
Psychotic	10	52	19	45	56
	11	68	24	54	12
	12	81	82	64	63
Mean		.610	.354	.603	.626

TABLE 27

ANALYSIS OF VARIANCE OF *Q*-SORT RELIABILITY COEFFICIENTS

Source	Sum of Squares	*df*	Variance
S type	640.5	3	213.5
Tests	6064.2	3	2021.4*
Tests × *S* type	4272.7	9	474.7
Within cells	16177.3	32	505.4
Total	27154.7	47	

* Significant at .05 level.

TABLE 28

REPORT-TYPE ITEM RELIABILITY COEFFICIENTS (TETRACHORIC CORRELATIONS) FOR THE 48 TJ's

Ss		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	72	67	80	91
	2	30	50	48	82
	3	73	59	93	73
Neurotic	4	92	79	88	93
	5	89	64	87	82
	6	85	52	92	83
Psychophysiological disorder	7	71	96	79	94
	8	74	58	94	90
	9	85	25	84	96
Psychotic	10	74	37	52	83
	11	90	00	59	61
	12	93	80	68	87
Mean		.773	.556	.770	.846

117 *True–False report-type items* and Table 29 the analysis of variance data. As in the *Q*-sort results, the average reliability figure for the TAT judges is significantly lower ($P = .01$) than that of the judges for the other three tests.

Table 30 presents the tetrachoric correlations between the answers given to the 100 *True–False factual items* in the first and second evaluations. The analysis of variance data (Table 31) indicates that the situation is exactly the same as for *Q* sorts and report type items, i.e., that TAT judges have a significantly lower ($P = .05$) reliability figure than judges of other tests.

TABLE 29

ANALYSIS OF VARIANCE OF REPORT-TYPE ITEM RELIABILITY COEFFICIENTS

Source	Sum of Squares	*df*	Variance
Tests	2383.6	3	794.5
S type	5648.8	3	1882.9*
Tests × *S* type	2484.1	9	276.0
Within cell	9766.7	32	305.2
Total	20283.2	47	

* Significant at .01 level.

TABLE 30

FACTUAL ITEM RELIABILITY COEFFICIENTS (TETRACHORIC CORRELATIONS) FOR THE 48 TJ's

Ss		Ror.	TAT	MAPS	MMPI
Non-psychiatric	1	70	86	88	83
	2	47	56	38	89
	3	75	60	60	60
Neurotic	4	87	67	75	73
	5	50	33	91	90
	6	86	66	83	81
Psychophysiological disorder	7	68	93	46	88
	8	83	47	89	84
	9	69	41	89	88
Psychotic	10	86	49	84	80
	11	77	47	89	48
	12	94	63	88	82
Mean		.743	.590	.767	.788

These reliability data may be summarized as follows:

1. Diagnoses based upon repeat evaluations of the same protocol vary from identical to markedly dissimilar.

2. Maladjustment ratings under the same conditions vary in a similar fashion with a mean absolute change of approximately 12% and a range from zero to 50%.

3. *Q*-sort reliability figures vary from −.07 to +.94 with an over-all mean of about .55 for all TJ's combined.

4. TAT judges make significantly less reliable *Q* sorts than judges of the Rorschach, MAPS, and MMPI.

TABLE 31

ANALYSIS OF VARIANCE OF FACTUAL ITEM RELIABILITY COEFFICIENTS

Source	Sum of Squares	*df*	Variance
Tests	333.1	3	110.4
S type	2912.9	3	971.0*
Tests × *S* type	2281.2	9	253.5
Within cell	8652.7	32	270.4
Total	14177.9	47	

* Significant at .05 level.

5. Report type True–False item reliability figures vary from zero to +.96 with an over-all mean for all TJ's combined of about .73.

6. TAT judges have a significantly lower mean personality type True–False item reliability figure than do judges of the other tests.

7. Factual True–False items reliability figures vary from .33 to .94 with an over-all mean for all TJ's of about .72.

8. TAT judges have significantly lower reliability figures for factual True-False items than do judges of the other three tests.

The outstanding general finding from the reliability data is that the reliabilities of the interpretations made from psychological tests vary considerably except for the individuals interpreting TAT materials (who were conspicuously low in their reliabilities). This variation of reliability cuts across the test used and the type of *S* for whom the protocol was interpreted. The most general implication of these findings has to do with the notion of the validity of the psychological tests, inasmuch as there are inextricable relationships between validity as defined and reliability of interpretation. Thus it would appear that efforts to establish the "validity" of our current psychological techniques—whether in terms of status validity, content validity, or congruent validity, etc.—are premature until adequate interpreter reliability has been established.

Discussion

Rather than recapitulate the results and the comments given in the preceding pages, the present section is limited to some final thoughts and opinions on the over-all implications of the study.

One of the authors has reported a survey on the status of studies concerning the validities of projective techniques (Little, 1957) which indicates that these studies, taken as a group, are contradictory or equivocal and in general do not give strong support to the contention that test procedures have clearly demonstrated validity. The general over-all results of the present study do not reverse this state of affairs. With a variety of types of *S*s, with more than one *S* in each category, with experienced judges doing blind interpretations in response to the same questions, the results of the analyses for four frequently used psychological tests did not yield the kind of positive correlations that would gladden the hearts of psychological test proponents.

The results are even somewhat more distressing when one considers the magnitude of correlations obtained even when they were significant. A small decrease in chance variation is a legitimate goal for screening instruments but relatively useless for tools such as projective techniques where individual prognosis is desired.

There are several legitimate objections to citing the results of the present study as indicating the level of clinical validity or reliability of the instruments used. "Validity" is a complicated and ambiguous term at best and the use of psychiatric consensus as a criterion does not improve it noticeably. The lack of agreement among the AJ's was quite clear for a number of the Test *S*s. Moreover, one may argue reasonably enough that judgments such as required by the tasks in this study cannot be made from a single information source and that validity can be determined only on batteries of tests as actually used in clinical situations. This assumption of a summative effect and/or an interactive effect of various instruments is a testable hypothesis, although it was not tested in the present study. An objection similar to this last one may be raised to citing the relatively low agreements among judges as evidence of unreliability of the instrument or the judge. Again the effect of combined information sources on inter-judge agreement is testable, but was not within the scope of this study.

In a certain sense, the study may have imposed on the TJ's a "naive clinical" requirement, to use Holt's (1958) schema. Blind analyses prevent the clinically desirable progressive checking and modification of hypotheses against other sources presumed to be more valid. However, such control was an essential element of the design of the study inasmuch as it was the *individual* instruments that were being assessed.

Recognizing these restrictions on the simple application of the results to tests as used in clinical settings, the results still indicate room for considerable improvement in test interpretation. For the tests used, agreement on diagnoses will be only slightly better than chance, judgments of maladjustment will be skewed toward the pathological, agreement with psychiatrists as to personality dynamics will be modest, and the clinician's reliability in all these areas will leave much to be desired. Since there were marked differences in the correlations among the different judges, perhaps each individual clinical psychologist and psychiatrist should make it part of his business to ascertain his own personal validity and reliability coefficient with various categories of patients, as measured against various criteria.

REFERENCES

Burton, A., & Harris, R. E. (Eds.). *Case histories in clinical and abnormal psychology.* Vol. 2. New York: Harper, 1956.

Friedman, I. Characteristics of the TAT heroes of normal, psychoneurotic and paranoid schizophrenic subjects. *J. proj. Tech.,* 1957, **21,** 372–376.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. *J. abnorm. soc. Psychol.,* 1958, **56,** 1–12.

Holtzman, W. H., & Sells, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.,* 1954, **49,** 485–498.

Little, K. B. Whither projective techniques: Paper read at California State Psychol. Ass., Los Angeles, March 1957. *J. proj. Tech.,* in press.

Little, K. B., & Shneidman, E. S. The validity of thematic projective technique interpretations. *J. Pers.,* 1955, **23,** 286–294.

Little, K. B., & Shneidman, E. S. The validity of MMPI interpretations. *J. consult, Psychol.,* 1954, **18,** 425–428. Also in G. S. Welsh & W. Grant Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. of Minnesota Press, 1956.

McNemar, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954.

Shneidman, E. S., Walther, J., & Little, K. B. *Thematic test analysis.* New York: Grune & Stratton, 1951.

(Accepted for publication May 17, 1958)

APPENDIX A

CASE HISTORY ABSTRACTS

TEST *S* 1

Identification. Subject 1 is a 25-yr.-old, divorced, Caucasian male veteran of English-Swedish descent. He is a "fairly religious" Methodist, i.e., he observes major religious holidays but does not attend weekly services; he feels that he conforms and adheres to all Protestant religious principles. He has worked at various semiskilled jobs: postal clerk, spray painter, sheet metal worker, salesman, post office, police officer. He is residing with his mother and father in a rented apartment in a large city on the West Coast. He was seen in a VA general hospital.

Description of patient's appearance. Patient is a neat, well-dressed, quiet, cooperative, and friendly young man who appears his stated age. His left arm and left leg are in a cast. He is neither tense nor manifestly anxious. Patient is 5 ft. 10 in. tall and weighs approximately 150 lbs. but looks smaller and lighter as he sits in his wheelchair. He presents the appearance of a "typical American boy" with clean-cut features. He speaks clearly without restraint and with no apparent emotional difficulties.

Presenting problem. The patient was involved in a motorcycle accident. He allegedly was not driving overly fast, but skidded on sand into the path of an oncoming automobile. He incurred a fractured left elbow and a fracture of the left tibia and fibula. He was treated for shock and received emergency fracture treatment and was then transferred to a large general hospital where he has received definitive treatment for his fractures. While in the hospital he presented no difficulties to the medical or nursing staff.

Psychiatric diagnosis. In the opinion of the psychiatrist, there is no psychiatric disorder. There are some acting-out trends in the character formation and a push for culturally accepted activities which could be presented by the patient to society as evidence of his strength and masculinity. However, to the present time these have not been so excessive as to be considered pathological. The patient's fear of inadequately playing the masculine role and his need to elevate his self-esteem and stature by constant activity and by mastering many things, plus the degree of restlessness he has evidenced in the occupational area, have not reached extremes which can be considered pathological.

TEST *S* 2

Identification. Subject 2 is a 23-yr.-old, single, Caucasian male veteran of Irish descent. He is a fairly religious Protestant who attends services about once monthly and on religious holidays. He is a student at agriculture school. He is residing with his parents, and previously lived in a dormitory at college. He was seen in a VA general hospital. He had entered a smaller hospital previously for treatment of multiple fractures following a motorcycle injury.

Description of patient's appearance. The patient is a pleasant, young adult sitting in a wheelchair with his right arm in a cast. His right ankle is slightly enlarged. The patient's right hip is fixed with a pin. He is in no pain or distress and appears happy and comfortable except for the limitations of the cast. Throughout all of the interviews the patient appeared cooperative and friendly. When erect he was approximately 5 ft. 8 in. tall, weighing approximately 140 lbs. He had a youthful, zestful expression and appeared no older than his chronological age. At no time did he show any obvious emotional conflicts or psychiatric difficulties. He had a crew hair cut. He had moderately strong features. He always spoke clearly without restraint and with no evidence of emotional tension. He stated that he intended to return to school after he had recovered from his injuries.

Presenting problem. The patient's entrance into the hospital was occasioned by physical injuries and multiple fractures incurred in a motorcycle accident which occurred when an automobile suddenly made a left turn without signaling. He was unconscious for a few minutes. An ambulance soon arrived which took the patient to the hospital. During the early days of hospitalization, he slept most of the time and had a groggy feeling during much of the time he was awake. He learned he had been operated on the day following the accident and that his right elbow and right hip and ankle were fractured. At the hospital he had relatively little pain, although chest pain from pneumothorax lasted for 4–5 days after his entry. The patient is ambulatory with crutches, but he spends most of the time in a wheelchair. He expects to leave the hospital in approximately 3 wks. He feels that he has relatively few problems about his physical condition and he is not concerned about the eventual outcome, having been informed that the bones were healing well.

Psychiatric diagnosis. In the opinion of the psychiatrist, there is no psychiatric disorder present. The personality trends shown by the patient, such as his acting-out tendencies, push into physical activity, push for culturally accepted activities, and the development of a "masculine approach" to adventurous and excitement-filled activities, have never appeared to an excessive or pathological degree.

TEST *S* 3

Identification. Subject 3 is a 23-yr.-old, single, Caucasian male veteran of Dutch descent. He is Protestant; not an active church goer, but attends church for principal religious festivities and prays when in stressful situations. He is a highly skilled radar repairman. He is residing with his mother and sister in their home. He was seen in a VA general hospital for minor surgery. He had an uneventful surgical experience and expects to leave the hospital within 2 wks. to return to home and work.

Description of patient's appearance. The patient appeared as a well-developed, well-nourished male who looked no older than his chronological age with youthful appearance and manners. He was approximately 6 ft. tall, weighing approximately 175 lbs. (His present weight is 167 lbs., patient having lost approximately 7 lbs. since his entry into the hospital.) Athletic habitus with no evidence of physical disproportion although patient's feet appeared somewhat small in comparison to his large bones and large physical frame. Patient had light blue eyes and reddish-blond hair. Facial features appeared regular and clean cut and patient gave a "typical American boy" appearance. He was very friendly, pleasant, and cooperative. He spoke clearly without showing physical evidence of emotional conflict and had no awareness of anxiety or emotional difficulties. He related well to the examiner. Ward behavior was reported as unremarkable on the nurses' charts and patient was considered friendly and cooperative by hospital personnel.

Presenting problem. Patient entered the hospital for a herniorrhaphy. Approximately 1 yr. ago, the patient noticed a small swelling in his right lower abdomen. He was checked by a physician in a nearby clinic and was told that he had a swollen gland in his groin. He had no further trouble with this until approximately 3 wks. ago when he noticed an increase in the swelling which now remained constant, whereas previously the swelling had fluctuated in appearance, disappearing for a few months and then reappearing. On this occasion, the patient noticed that the swelling was constant, was larger, and appeared lower down in the abdomen. He himself recognized this as a hernia and was able to reduce the swelling by pressure. He visited his local physician who confirmed the patient's impression and recommended that the patient enter the hospital for this elective surgery. Patient has no other physical difficulties.

Psychiatric diagnosis. In the opinion of the psychiatrist, there is normal personality structure and the absence of any definite psychopathology in this case. There is evidence of ability to adapt to and adjust to life's problems and experiences. It is considered that the patient falls within the normal personality group without definite psychiatric disorder.

TEST *S* 4

Identification. Subject 4 is a 27-yr.-old, married, Caucasian male veteran of English-Dutch descent. He is a Protestant and very active in religious observances. He is a salesman by occupation. He is residing with his wife and children in their small home. He was seen at a VA neuropsychiatric hospital.

Description of patient's appearance. Patient is a 5 ft. 10½-in. tall, 170-lb., pleasant-appearing male with wavy brown hair, blue eyes, and a freckled youthful face. He presents the appearance of a "typical American farm boy," frank in demeanor and very direct and straightforward. As he speaks there is a suggestion of a stammer and he appears slightly ill at ease most of the time. Early in the interviews he spoke slowly and with a regular, even tone. Occasionally there is fine tremor of the hands. Toward the end of the interviews he appeared much more at ease but still continued to show slight signs of discomfiture. He is always pleasant, cooperative, and friendly, and relates very well to the examiner. Although during the interviews he always appears calm and peaceful superficially, there is evidence of underlying tension and at times a suggestion of considerable hidden anger. The patient appears moderately neat, well dressed, and somewhat younger than his chronological age. His clothes give the appearance of a relatively recent weight gain, being generally tight and close-fitting. The patient perspires rather freely. He began to express his upset feelings and feelings of anger in particular as his relationship with the examiner improved. Except for his anxiety and tension the patient showed no other psychopathological features during the interview. He was always fully oriented and in contact with the examiner and his environment. No frankly bizarre signs or symptoms were ever in evidence. The patient appears of superior intelligence, shows no signs of intellectual deterioration and gives no evidence of psychopathological phenomena during the period of observation.

Presenting problem. The patient was transferred to the neuropsychiatric hospital from a large county hospital psychopathic ward after he had had a confused episode with amnesia. A year before, he had an episode of acute anxiety with depression and short periods of "unconsciousness" and was hospitalized at the county hospital for 12 hr. with a spontaneous remission at that time. The patient has been receiving outpatient psychotherapy at a mental hygiene clinic since that time. Immediately prior to his precipitating emotional disturbance the patient had been drinking considerably and also had been taking barbiturates. The episode of confusion and amnesia was self-limited, remitting in approximately one day and the patient was referred for hospitalization as a result of the impression of the psychiatrist conducting the outpatient psychotherapy. He felt the patient was not benefitting materially from his outpatient psychiatric care. In view of the patient's poor progress on the outpatient level it was recommended that he enter the NP hospital. Although the wife filed commitment charges, the patient entered voluntarily for psychiatric treatment. The patient has shown signs of mild anxiety, increasing irritability, and at times mild confusion with reduced work performance, and he has had previous episodes of what were considered to be hysterical seizures since his discharge from the service at the age of 21, six years ago.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of a chronic anxiety reaction manifested by mild to moderate despondency and depression. In the acute anxiety state the patient developed an hysterical conversion psychoneurosis, manifested by acute confusion, dissociation, emotional lability, free expression of his underlying aggressive feelings, and brief episodes of amnesia with loss of self-identity and self-misidentification. Also in the past there has been utilization of an "hysterical" technique of conversion paralysis.

TEST *S* 5

Identification. Subject 5 is a 29-yr.-old, married, Caucasian male veteran of Scotch-Dutch descent. He is Protestant; attends Sunday services fairly regularly and prays nightly although he is somewhat embarrassed about the adherence to nightly prayers. He has sentimental feelings about major religious holidays which he observes regularly. He does not feel himself to be rigid or orthodox in religious beliefs and principles. He is a mechanic and previously had various semiskilled jobs. He is residing with his wife and the wife's child in a home purchased by the wife. Patient was threatened by his employer with forced resignation unless he obtained treatment for his nervous state. He was pressured to take one year's leave of absence for observation and psychiatric treatment. He was seen at a VA neuropsychiatric hospital.

Description of patient's appearance. Patient is a well-groomed, well-developed man, 6 ft. tall, and weighing 200 lbs. He has brown hair and brown eyes and a neatly trimmed mustache. He appears cooperative and friendly. He has a husky, broad-shouldered appearance, appears his chronological age, and always is cheerful, friendly, and cooperative, with a somewhat ingratiating manner. At all times he was well groomed and dressed carefully. In all interviews he was quiet and showed no very intense emotional disturbance, but constantly showed presence of an underlying tension state. This was manifested by tremulousness, mild perspiration, and at times mild pressure of speech, mild embarrassment when discussing his recent history of emotional disturbances, and mild increase in motor activity as the interviews progressed. At no time was there any bizarre behavior or abnormal thought content, and the patient showed no evidence of psychotic behavior or ideation. His affective state did not appear remarkable except for the underlying tension. The emotional lability which he described was not evidenced in the interviews.

Presenting problem. The patient served with the Army Air Force for two years. After a number of missions as a radioman he was returned to the United States where his anxiety reaction became manifest and disabling for the first time. After he had volunteered for an additional tour of duty he was re-examined and discharged with a diagnosis of psychoneurosis, anxiety state, and was given a pension for service-connected disability. Since then there has been increasing "nervousness" manifested by high peaks of inattention, marked tremulousness of the hands to a degree where the patient is unable to work or feed himself, marked irritability, anger outbursts, emotional lability with occasional uncontrolled crying spells, fears of losing control of his feelings, and fears of appearing insane. He also had pressure headaches, insomnia, nightmares, increasing withdrawal from friends and acquaintances, and concern and irritability over trivial, inconsequential matters. Marked irritability with family members, in particular his wife and children, also exists. He has occasional dizzy spells and "blackouts" usually associated with his drinking. During such "blackouts" the patient has acted aggressively and antisocially. The patient has had sufficient marital friction and dissatisfaction with his marriage to cause him to divorce two wives since his discharge 7 yr. ago, and he is now unhappily married to his third wife, 10 yr. his senior. In the past 6 mo. the patient has complained of increasing gastro-intestinal disorders with indigestion, heartburn, belching. There also has been increasing frequency of "blackouts" and dizziness which occur when he "looks down or moves about rapidly and changes from a sitting to a standing position rapidly." At times he has noted that objects in front of him become indistinct and even completely disappear. Nail biting, insomnia, headaches, night sweats, and nightmares have become more frequent and have interfered markedly with his interpersonal, occupational, social, and marital adjustment. The patient is embarrassed and ashamed about his overt nervousness and tremulousness which he is unable to control, and he shows marked irritability and belligerency when attention is called to his physical tension state. He has become involved for the past few years with civilian authorities as a result of antisocial, dangerous behavior committed while under the influence of alcohol, but the patient denies that he is an alcoholic. He recognizes, however, that he drinks heavily when his tension state is high. The patient traveled to the city where the hospital is located alone and went on a drinking spree immediately prior to admission, during which time he lost his wedding ring and wallet. Since his entry into the neuropsychiatric hospital there has been a marked reduction in his anxiety reaction.

Psychiatric diagnosis. This case presents the psychiatric features of (*a*) an anxiety reaction, moderately severe, manifested by marked irritability, inner tension state, insomnia, night sweats, nightmares, headaches, excessive perspiration, gastro-intestinal distress, occasional syncopy, and feelings of inadequacy and inferiority; (*b*) a motor tension state manifested by moderate to marked tremulousness; (*c*) symptomatic alcoholism; (*d*) aggressive hostility, and occasional antisocial behavior—in an emotionally immature and unstable personality. In addition, the patient showed mild to moderate obsessive trends which are manifested as perfectionistic tendencies and a desire for orderliness and cleanliness and paranoid features seen as feelings of social isolation, mild ideas of reference, and feelings of discrimination.

TEST *S* 6

Identification. Subject 6 is a 33-yr.-old, married, Caucasian male veteran of Swedish descent. He is a Protestant, not orthodox in religious attitudes or observances. He is floor manager in the sporting goods section in a large department store. He resides with his wife and 4-yr.-old child in a rented apartment and has been in a general hospital for 3½ mo. prior to transfer to the present psychiatric hospital for further observation and treatment. He came into the VA neuropsychiatric hospital as a voluntary admission.

Description of patient's appearance. The patient is a healthy appearing, well-developed, and well-nourished adult, 5 ft. 10 in. tall and approximately 170 lbs. in weight. He is clean shaven with regular, strong features and a strong, rugged appearance. He has a ruddy complexion with brown hair and brown eyes. Hands and chest appear hairy. Although the patient was always pleasant and cooperative, there was most frequently an undertone of mild depression, at times mild anxiety, manifested by some sighing and hyperventilation. Typically the patient presented himself as a controlled, slightly tense individual with tension showing in his facial expression, bodily movements, and gestures. There was an excessive amount of control shown in his speech and his movement. Speech was ponderous and controlled without appearing pedantic. He appeared of superior intelligence and had more than average sophistication in psychiatry. His affective state was markedly controlled and always appropriate to the circumstance. At times when speaking about his relationship with his wife and his recent psychiatric difficulties he became slightly tremulous and more tense. No frankly compulsive behavior was observed during the interview. Occasionally there was mild twitching of the facial musculature. Although the patient was always overtly friendly and cooperative toward the examiner, at times he expressed antagonism and hostility toward the hospital authorities with expressions of resentment at the treatment accorded him. The patient was always early for his appointments and appeared sincerely interested in meeting with the examiner in order to obtain help. The patient had no physical complaints.

Presenting problem. Because the patient's previous response to psychotherapy at a general hospital was poor, he was transferred to a psychiatric hospital. His chief psychiatric complaints had their onset allegedly while he was in the service eight years prior (in 1944). They have markedly increased in the last few years until the patient has become occupationally and socially disabled, estranged from his wife and family, and much more aware of his inner tension state. There have been progressively increased intermittent exacerbations and remissions. In the last year the symptoms have markedly increased in severity and frequency. The patient receives a veteran's pension for his neuropsychiatric condition—head tremor, anxiety reaction. Also, in the last few years there has been a development of hand tremor with progressive difficulty in manipulating and controlling his hand movements. A second complaint, numerous obsessive and compulsive trends which are not considered really disabling by the patient. These consist of an excessive need for cleanliness, fear of dirt, overconcern with human dirt and excreta, hand washing, compulsive needs to be certain that doors are locked, cigarettes in ashtrays put out, and obsessive counting of cracks in sidewalks, counting floors in buildings, and other compulsive and obsessive reactions. Thirdly, he has phobic reactions of a mild degree consisting of fear of crowds, fear of losing control of himself and of showing his emotion which he attempts to control by exaggerated speech mannerisms, etc. Fourth, the patient experiences a sub-acute anxiety state manifested by over tension and restlessness, mild facial grimacing, and gripping of hands and extremities. Fifth, there is an increasing withdrawal from his friends and family with daydreaming increasing in intensity and frequency. The patient shows considerable self-reproach, self-criticism, and loss of self-esteem with the subjective experience of mild to moderately acute depression. A tension state accompanies the depressed mood. The patient has a moderate degree of insomnia, mild loss of appetite without loss of weight since the onset of his psychiatric difficulties. All of the patient's symptoms allegedly increased markedly in the past year prior to his entry into the hospital. Although there has been a decrease in the psychiatric signs, the patient's inner feeling of upset and turmoil have continued unabated and the depressed mood has increased since his entry into the hospital.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of a mixed psychoneurosis in a schizoid character with obsessive-compulsive trends. The neurosis is manifested by anxiety symptoms, tremor of head and hands, and inner tension state, phobic reactions, obsessive-compulsive reactions, strong depreciatory trends with guilt feelings and a mild to moderate depression. There is also marked social withdrawal from group and immediate family relationships.

TEST *S* 7

Identification. Subject 7 is a 31-yr.-old, married, Caucasian male veteran of English-Scotch-Dutch descent. He is a Protestant, religious in attitude, but not orthodox in observance. He is a waiter, but aspires to be a musical composer. He is residing with his wife in a rented apartment. He was seen at a VA general hospital.

Description of patient's appearance. Patient is a neat, well-dressed, young adult who appears no older than his chronological age. He is tense, and shows restrained appearance and mannerisms. He speaks with a low, modulated, restrained voice, without moving lips very much or opening mouth wide. No evidence of emotion during session except for patient's gratitude at opportunity for talking with examiner. Patient was late for almost every session and openly very apologetic on this account. He was friendly and overtly cooperative and pleasant.

Presenting problem. Chief complaint is presence of generalized sharp, diffuse, colicky abdominal pains with alternating diarrhea and constipation since military service in 1944. He was diagnosed as "irritable colon syndrome." Physical examination, laboratory tests and X rays all negative for organic pathology. Since 1944, patient aware of "nervousness" (anxiety and tension). Exacerbation and intensification of complaints in past year (1951).

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of an anxiety reaction chronic, mild to moderate, with somatization reactions of a gastro-intestinal "irritable colon syndrome," accompanying the anxiety reaction and tension state.

TEST *S* 8

Identification. Subject 8 is a 29-yr.-old, single Caucasian male veteran of English-Irish descent. He is Protestant. He is a carpenter and is residing with his parents in their home. He was seen at a VA general hospital.

Description of patient's appearance. The patient is a neatly dressed, well-developed and well-nourished, mildly obese, young white adult who appears a few years younger than his chronological age. He is 5 ft. 11 in. tall and weighs approximately 200 lbs. He has a general florid and ruddy appearance which allegedly resembles that of his father. In facial appearance, however, he says he resembles his mother. During the first interviews the patient was slightly tense but soon appeared to feel at ease with the examiner. In his speech and behavior during the initial interview he appeared shy, restrained, and inhibited and this restraint continued to some degree throughout all of the interviews in the discussion of his personal life. He appeared fearful of offending the examiner but with increasing contact became much more friendly and demonstrative, always holding back somewhat as if "on guard." He always appeared early for his interviews and appeared most eager to recount his personal history with the examiner. During most of the time he was interviewed he was awaiting transfer to the surgical service for peptic ulcer surgery.

Presenting problem. This is the patient's first admission to this hospital for treatment of a bleeding peptic ulcer. Abdominal complaints have been present since 1956 at which time a diagnosis of pyloric spasm without ulcer was made on X ray. The patient has a history of two previous episodes of gastro-intestinal bleeding in 1947 and 1948 and has had one admission to another general hospital in 1949 with a diagnosis of a perforated ulcer. Prior medical and conservative dietary therapy have produced transient relief. On this admission the patient entered the hospital when he recognized from his black, tarry stools and his increasing feelings of weakness that he was again experiencing a gastro-intestinal hemorrhage. Although the patient has associated his previous bouts with emotional disturbances, he offers no knowledge of any external precipitating factors of emotional disturbance which can be related to the present physical upset.

Psychiatric diagnosis. In the opinion of the psychiatrist this case presents the psychiatric features of a mild anxiety reaction, chronic, with a mild degree of depression. The primary psychiatric diagnosis would be a psychophysiologic gastro-intestinal reaction with peptic ulcer syndrome and sequelae.

TEST *S* 9

Identification. Subject 9 is a 26-yr.old, single, Caucasian male veteran of French and English descent. He is a nondenominational Protestant and is not actively religious. He works as a machinist in the aircraft industry. He resides with his mother in the mother's home. He was seen at a VA general hospital.

Description of patient's appearance. The patient is a tall, moderately well-developed, nourished man with asthenic habitus, appearing thinner than his actual weight of 155 lbs. He is also somewhat ungainly. He looks a few years older than his chronological age. He is dark complexioned with dark hair, is neatly dressed and appears pleasant, friendly, and cooperative. During all of the interviews he appeared superficially relaxed but was evasive in discussing his family and personal history. Although at first it was believed that he was deliberately and consciously evasive, the examiner's impression terminally was that the patient was markedly naive and ignorant about himself, his feelings, and feeling-experiences. He presented a picture of naiveté and ignorance of his personal experience which was accompanied by a marked passivity, a blandness and lack of concern over any problems at home or with his fiancée. The patient's lack of introspection and lack of subjectivity was so marked that he could tell the examiner very little about himself because he remembered so little. He appeared to have repressed more than suppressed much of his personal history. When he was led to discuss more meaningful material he became more restrained and even more evasive, although he was unable to recognize his evasions. When displaying his feelings he showed facial mannerisms and exaggerated surprise, amazement, and approval. On two occasions, when material about sexual relations, masturbation, and his feelings about his mother were brought up, the patient showed some objective anxiety and a flare-up of awareness of uneasiness and inner tension. At these times he immediately avoided further discussion of the material which appeared to disturb him. He showed a complete inability to express directly any hostility, dislike, or even unfavorable criticism. Pleasant and casual situations were stressed by him and he apologized for his inability to "like" everybody.

Presenting problem. The patient entered the hospital with a diagnosis of atopic dermatitis. He was well until early 1952 when he developed an intermittent pruritus (itching) on the back of his neck and the extensor surfaces of both hands. Approximately 1 mo. later a gradually spreading rash developed which has at times involved the larger portions of his face, his upper extremities, especially on the flexor surfaces, and parts of his trunk, thighs, and lower extremities. The patient has been treated by a private physician and has responded temporarily to cortisone. However, when he was taken off cortisone the dermatitis had an exacerbation. For 5 mo. there were remissions and exacerbations that followed no consistent pattern according to the patient. Although he was again placed on cortisone he did not respond as previously and thus sought hospitalization. The patient previously also had X-ray treatments and various topical aids to which he did not respond. In the hospital the patient showed moderate improvement with conservative therapy but showed a marked exacerbation of his dermatitis each time he left the hospital to go home to visit his mother and fiancée over the week end. Because of this specific temporal relationship, the dermatologist referred the patient for psychosomatic evaluation.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case is diagnosed as a psychophysiologic skin reaction manifested by atopic dermatitis, present in a passive-dependent character with marked fear of, and inhibition in, expressing his anger and sexual desires.

TEST *S* 10

Identification. Subject 10 is a 22-yr.-old, single, Caucasian male veteran of Dutch-Scotch-English descent. He is Protestant and religious in attitude and observances. He is a student. He resides with his mother and stepfather at home. He was self-referred into the VA neuropsychiatric hospital, where he was seen.

Description of patient's appearance. The patient is a well-developed and well-nourished, young male who appears somewhat older than his stated age. He has brownish-red, blonde hair, a ruddy complexion, and blue eyes. He is about 5 ft. 6 in. tall and weighs approximately 150 lbs., with the general appearance of a stocky, athletic build. He is left handed, has small features and small hands. He appears friendly, cooperative, and pleasant toward examiner, but he is extremely anxious and tense and at times mildly agitated. He has a moderately fine tremor of the hands. His face flushes easily and he perspires readily, especially about the hands and forehead. At times during the interview he broke down tearfully. He spoke in a very low monotone. His conversation was connected and coherent except on a few occasions when there was marked pressure and tension. At such times there was mild disconnection. He showed no overt hostility. Often he appeared depressed and at times definitely suspicious. At no time during the interviews or during the period of observation on the ward did he show active hallucinations. He appeared fully oriented and showed no obvious memory defects. Toward the end of his hospital stay the patient noted that he was concerned about some of the other patients who had recently been admitted because he felt they were looking at him and spying upon him. He connected his ideas of reference and mental turmoil with various emotionally disturbing experiences he had experienced, and also to disturbing relationships with his relatives, his mother in particular. There was partial awareness of the seriousness of his mental condition and a desire for treatment. This awareness of inner upset was markedly reduced as he became increasingly more disturbed. The patient was seen approximately 1 wk. before admission to the hospital by a physician who recommended that he enter the neuropsychiatric hospital as a voluntary patient. Although the patient was interested and eager for this treatment, both his mother and stepfather were strongly against his hospitalization and also against his receiving the somatic treatment which was subsequently recommended by the hospital staff.

Presenting problem. Although the patient's premilitary service adjustment showed no frankly bizarre features, he appeared to be somewhat withdrawn at that time and showed a rather schizoid adjustment. At the age of 20, prior to his entry into service, he had already given evidence of mild confusion, irritability, restlessness, and inability to plan for the future. He enlisted in 1950 and for the first 6 mo. appeared to make a satisfactory adjustment. From the beginning of his overseas duty in 1951 he began to show evidence of increasing confusion and had bizarre personal experiences. He felt that he was subjected to excessive stress and undue responsibility in his military duty. Feelings of unreality developed. He was struck by the "awesome" beauty of the scenery and felt that he had been there previously, had seen the surroundings and also the native inhabitants. He became frightened by these feelings and by a premonition of events which, according to the patient, actually occurred subsequently. He felt somehow that he had influenced these events and experiences. Following a homosexual experience in 1951 and the NP hospitalization of his best friend, whom the patient had recognized was frankly psychotic (he had had hallucinations and bizarre and frightening ideas), the patient became more confused and frightened. He developed marked headaches, ringing in the ears, depression, fears of suicide, increased feelings of anger and resentment at the position which he held in the military, difficulties in thinking and concentrating, and a marked increase in nervousness and irritability. He was troubled with many conflicting ideas and was in a constant turmoil because of the inability to commit himself to any definite action. He became ruminative about his actions to the point of being unable to act. He began to think people were calling him names and looking at him, especially in crowds. He developed insomnia and lost weight. He entered a military hospital for a period of 1 mo. and then returned to active duty. In early 1952 he returned to the hospital with an exacerbation of his symptoms. He became suspicious of those about him, very anxious and tense, now doubting his sanity, and much more depressed. He was given a medical discharge from military service later in 1952. The patient moved west to live with his mother and became increasingly more upset as a result of her difficulties with the stepfather. The anxiety increased, he developed extreme awareness of his heartbeat, palpitations, and chest pains. He felt light headed and felt as if he were constantly febrile. The headaches became more severe and the patient sought treatment, upon which he was referred to the psychiatric hospital.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of a schizophrenic reaction, paranoid type, following the early acute undifferentiated type of schizophrenic reaction. The patient first showed confusion of thinking and emotional turmoil with perplexity, ideas of reference with weird and peculiar ideas with feelings of awe and premonitions. Auditory hallucinations at first were mild and unclear and never developed into very clear-cut hallucinatory phenomena even in the more fully developed paranoid state. The outstanding psychiatric condition early developed was the perplexity and emotional turmoil with increased ambivalent feelings and confusion of thinking. Subsequently the thinking became much more unrealistic and autistic with a more definite development of delusions of persecution and ideas of reference. The behavior became more impulsive and unpredictable and the underlying anger, resentment, and aggressive feelings were obvious. Also there developed the somatic symptoms of headache and scattered physical pains, in particular the chest pains referred to in the heart area. An anxiety and tension state continued strong throughout the months of development of the schizophrenic reaction. The loss of contact with reality and the development of the bazarre symptomatology are the most clear-cut symptoms.

TEST *S* 11

Identification. Subject 11 is a 29-yr-old, single, Caucasian male veteran of English descent. He is a Protestant and very religious in attitude and observances, with marked conflict about this. He has done various types of unskilled labor. He resides most frequently with his father. He was seen at a VA neuropsychiatric hospital at the time of his second admission.

Description of patient's appearance. The patient is a tall (5 ft. 11 in.), neatly dressed, moderately obese, male with a florid facies, who appears tense, anxious, and perspires profusely. His face is immature and boyish. He blushes easily, appears earnest, cooperative, friendly, and pleasant, most amiable and most eager to please. The patient superficially gives the impression of being dull, simple, and phlegmatic, but more intensive contacts leave the definite impression of at least normal intelligence which is reduced by a pathological process. He appears blundering and gross, naive and childlike, sensitive and easily hurt by criticism. His affect is blunted but not inappropriate to the situation or to the context of conversation. He was fully oriented and conversation was clear, connected, and coherent. No overt psychotic features appeared present at this time although they were reported in the history and by the patient. There was no obvious memory or intellectual defect and the patient had "partial insight" into his alleged hallucinatory experiences; his present judgment appeared fairly good. He appeared aware of his difficulties in adjusting socially, economically, and occupationally outside the hospital. He himself openly sought the protection of institutional life at this time. He recognized his social disability and limitations and was able to tolerate recent criticisms and joking directed toward him from fellow patients much better than he had been able to tolerate such criticism prior to hospitalization. He felt much more at ease in the hospital than he did outside. He was openly eager for approval and acceptance with a simple undisguised fear of rejection. He appeared to establish without any difficulty a good relationship with the present examiner.

Presenting problem. The patient was first admitted to the neuropsychiatric hospital in 1945 following his discharge from the service with a diagnosis of psychosis with psychopathic personality; psychosis in remission manifested by nomadism, depression, confusion, irritability, and religious conflicts. He left the hospital 2 mo. later, after receiving supportive treatment with custodial care and showing some improvement by being less confused and less depressed. Subsequently he lived a lonely, isolated social and nomadic life with borderline social and economic adjustment, frequent job changes, being fired often because of his poor work. He lived alone in the southeastern portion of the United States for the following year. In 1946 he moved to the northwest to live with his father, a preacher. Here he held occasional jobs and also had very limited social contacts in church. Soon after his discharge from the hospital the patient began to gain weight and in the next year he gained approximately 100 lbs. In 1947, following a VA pension examination, the patient's diagnosis was changed to a schizophrenic reaction, unclassified type. The patient continued to have strong differences of opinion with his father who, because of his religious beliefs, objected strenuously to the patient's occasional smoking, card playing, and occasional drinking and dancing. As a result of these continuing conflicts, in 1952 the patient left his father as he had done frequently before for similar reasons. He became increasingly more nervous and hyper-irritable (especially sensitive to loud noises), and had a return of gastric distress with belching after meals and chest pains, with substernal pressure symptoms, in addition to becoming increasingly weak and tired. Also, in the past year the patient had severe generalized headaches, slept poorly, had nightmares, and had a return of ideas of reference and auditory hallucinations. He heard passers-by on the street whispering unkindly comments about his physical appearance and excessive weight, criticizing his unemployed status, being foolish, and having general difficulties in adjustment. Also, he heard "The Lord" telling him that he was sinful, not a good man, and not a good Christian. The patient was aware of his religious preoccupations which he felt were unhealthy and also aware that the voices and ideas of reference were related to his confusion and emotional upset. He became depressed and confused, felt lonely and began blaming himself for his failures. He felt that he had foolishly squandered the little money he had, felt frustrated socially, occupationally, and sexually, and unhappy about his occasional arrests for vagrancy. He was aware of his increasing difficulties in adjustment and entered the neuropsychiatric hospital for the second time in 1952 as a voluntary admission in order to gain the protection of institutional life. When examined, the hallucinations and ideas of reference no longer were present and the patient had adjusted very well to the neuropsychiatric ward. However, he was fearful of a return of his disturbing symptoms following his discharge from the hospital and was concerned about his future adjustment.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of a schizophrenic reaction, chronic, undifferentiated, with periodic exacerbations of acute confusion with ideas of reference and auditory hallucinations. Also there is the development of somatization reactions in the gastro-intestinal and cardiovascular systems and general physiologic reactions accompanying the anxiety and tension states.

TEST *S* 12

Identification. Subject 12 is a 25-yr.-old, single, Caucasian male veteran of Danish descent. He is Protestant and is religious in attitude, observances, and approach. He is unemployed. He resides with his mother and father in their home. He entered a VA neuropsychiatric hospital as a voluntary admission.

Description of patient's appearance. The patient is a moderately well-developed, thin, fairly well-nourished, adult white male of asthenic habitus, who is poorly groomed and needs a shave. During the interview he appeared preoccupied and at times lost contact with the examiner. He stared off into space and appeared lost in his own thoughts. He sat rigidly in his chair with a very tense expression. His right hand was clenched into a fist with the nails digging deep into the palm. The patient refused to open his hand voluntarily but this was done with gradual pressure. When his limbs were moved about they retained their new position for short periods of time and then gradually returned to their old position. The patient rarely looked at the examiner and occasionally made facial grimaces which appeared unrelated to the material being discussed. He answered questions briefly but logically and coherently. There was considerable denial, evasion, and at times some circumstantiality. Especially when discussing his relationship to his father did the patient appear evasive. He spoke very slowly and frequently would clamp his teeth together for no apparent reason. During the interview he frequently picked at his fingers and nails. Frequently he would squirm and twist during the interview and showed considerable difficulty in talking. There were frequent pauses. The patient showed marked blunting in his affect and psychomotor retardation as well as physical retardation. He spoke almost entirely in a monotone and at no time were there changes in inflection. He appeared oriented as to time, place, and person and denied the presence of hallucinations and delusions upon direct questioning. His preoccupied appearance, bewilderment, and perplexity, however, led to the clinical impression of hallucinatory phenomena. He was sufficiently in contact to say at the end of one interview that he realized that he was uncooperative and that he hoped "things would be better next time." On the ward the patient was uncommunicative, evasive, and uninformative. He showed marked compulsive behavior and had a constant habit of punching his fist against various objects. Much of the time he appeared to be out of contact. However, when pressed he could be made to interact somewhat with the examiner.

Presenting problem. The patient stated that he had no complaints and alleged that he was in the hospital to obtain a rest and to improve his health and strength. However, investigation revealed a gradually increased withdrawal from family and friends which had been noted since 1947 when he had incurred an injury (on his head) and was subsequently discharged from military service. Further investigation, however, revealed that the patient had been nervous and a management problem since early childhood. The withdrawal, seclusiveness, and poor adjustment was evidenced in military service prior to his injury and had only become exaggerated following the injury and discharge from service. Since 1947 the patient's psychomotor and physical retardation progressed to the point of occasional periods of stupor accompanied by auditory hallucination and delusions. Inappropriate laughter and impulsive rage unrelated to the immediate circumstances had been present since 1951. The patient has had psychiatric treatment at a mental hygiene clinic and also by a private psychiatrist. There have been fluctuations in his clinical course but there has been gradually increasing severity of his psychiatric symptoms in the past few years. Although the patient entered the hospital voluntarily, he was pressured into the hospitalization by his mother who threatened commitment unless he agreed to coming to the hospital.

Psychiatric diagnosis. In the opinion of the psychiatrist, this case presents the psychiatric features of a schizophrenic reaction primarily catatonic in type, severe, manifested by auditory and visual hallucinations, unsystematized delusions, bizarre and obsessive-compulsive behavior, catatonic withdrawal with psychomotor retardation, and other rather typical catatonic features.

APPENDIX B

ALPHABETICAL LISTS OF THE JUDGES

TEST JUDGES

Anderson, Dorothy
Armon, Virginia
Aumack, Frank L.
Baker, Gertrude
Barron, Francis X.
Borenstein, Betty A.
Brackbill, Glen A.
Bugental, James F. T.
Crook, G. Hamilton
Dahlstrom, Grant W.
Davenport, Beverly F.
Deigh, Maurice
Farberow, Norman L.
Fine, Reuben
Fisher, Jerome
Forer, Bertram R.
Goldenberg, Herbert
Gough, Harrison G.
Gross, Zoltan
Guthrie, George M.
Halpern, Florence C.
Hamster, Richard C.
Hanvik, Leo J.
Harrow, Gertrude S.
Hohmann, George W.
Hubbard, Ruth M.
Jacobs, Alfred
Klopfer, Walter G.
Kutash, Samuel
Little, Jack F.
McDonald, Franklin R.
Marsh, James T.
Miller, Christine
Olinger, Leonard B.
Patterson, Virginia
Piotrowski, Zygmunt
Proud, Ann P.
Ross, Harvey L.
Schofield, William
Shapiro, David
Spiegelman, J. Marvin
Stauffacher, James C.
Thompson, Clare W.
Vineberg, Shalom E.
Walker, Robert G.
Welsh, George S.
Wesley, S. Medford
Wisham, Wayne

ANAMNESTIC JUDGES

Brill, Norman Q.
Conrad, Simon J.
Coodley, Alfred
Firestone, Milton
Frostig, Jacob Peter
Futterman, Samuel
Gabe, Sigmund
Gonda, Thomas A.
Goodstone, Gerald L.
Ingham, Harrington V.
Jacobs, James S. L.
Lester, Milton
Meyer, Mortimer M.
Morgan, David W.
Nissen, Archie
Pumpian-Mindlin, Eugene
Rangell, Leo
Saperstein, Jerome L.
Schreiber, Manuel
Seidman, Julius
Sturdevant, Charles
Unger, William
Wagner, Phillip
Walker, E. F.

APPENDIX C

ITEMS OF TASKS 2–5

SCALE OF ADJUSTMENT

Instructions: Indicate where you believe this subject fits on the scale below by putting an "X" at the appropriate place on the vertical line.

Score[a]	
0	Making a good or adequate adjustment.
1	
2	Functioning adequately, but at considerable psychological cost.
3	
4	Making a marginal or precarious adjustment.
5	
6	Poor adjustment; should have psychiatric care or be in a hospital.
7	
8	Is a custodial case; cannot survive without care.

[a] This column was not present on the rating scale. It is presented here to indicate how scores were assigned to equal divisions of the vertical line.

REPORT TYPE TRUE–FALSE ITEMS

Instructions: Items 1–118 are statements (or inferences) about the subject's personality. On the basis of your contact with the subject, answer each item by circling T (True), F (False), or ? (Cannot say) on the Answer Sheet. Use ? (Cannot say) as infrequently as possible.

1. Lack of emotional security is his major problem.
2. Rarely expresses any strong emotion.
3. Is in poor contact with reality.
4. Feels that he is not worthy of the love of others.
5. Has at least bright normal intelligence.
6. Indulges in counterphobic masculine activity.
7. His fear of retaliation blocks any overt expression of his hostility.
8. Possesses much drive and ambition.
9. Few things seem to worry him much.
10. Is a fast thinker.
11. Feels, unrealistically, that he is under the influence of others.
12. He feels, deep down, that he is a no good, not worth while person.
    (This item was inadvertantly omitted from the judges' lists.)
13. Considers sex a rather shameful thing.
14. His thinking and ideas tend to be clear-cut.
15. His sexual needs are only moderate.
16. Has marked independence strivings.
17. Has a great need to be loved.
18. Is extremely self-conscious.
19. Is often hostile without being aware of it.
20. Generally turns to alcohol for release when he is tense or anxious.
21. Is quite mature intellectually.
22. Shows a strong tendency to intellectualization.
23. Is a calm person.
24. Heterosexual relations frighten him.
25. Is extremely competitive.
26. Needs affection but in a mature relationship.
27. Shows relatively few objective signs of anxiety.
28. Has considerable hostility toward authority figures.
29. Gets along better with men than women.
30. Generally develops some physical symptom when tense or anxious.
31. Is demanding and self-centered.
32. Becomes depressed very easily.
33. Is sensitive and introspective.
34. Shows no great fear of rejection.
35. His intellectual approach to problems is logical and orderly.
36. Tends to blame his mistakes upon others.
37. His behavior is somewhat effeminate.
38. Is neat and orderly.

39. Is frustrated by his inability to live up to his own high standards.
40. Has little need for affection.
41. Gets angry usually only when appropriate.
42. Is quite hostile towards women.
43. Some of his complaints resemble bodily delusions.
44. Sees most of his peer group as competitors.
45. Has episodes of marked emotionality.
46. Seems to accept his hostile impulses without great conflict.
47. Has considerable fear of getting emotionally involved.
48. Is a dominating individual.
49. Is fairly rigid and inflexible.
50. Is somewhat overly logical and precise.
51. His thinking is quite unrealistic.
52. Considers himself inferior to others in most respects.
53. Has a number of rather rigidly moralistic attitudes.
54. Is a mild and self-effacing person.
55. Has a definite homosexual conflict.
56. Is below average in intelligence.
57. Considers most people he knows as inferior to himself.
58. Has a reasonable amount of confidence in his own ability.
59. Plans things fairly effectively.
60. Seldom feels depressed.
61. His strong aggressive impulses are a source of marked conflict for him.
62. Has an infantile need for affection.
63. Is a restless and impatient person.
64. Has a strong desire to rebel against authority figures.
65. Has achieved a comfortable separation from his parents.
66. Is a cold and unresponsive person.
67. Feels adequate to handle most problems he meets.
68. His thought processes are confused.
69. Is rather unimaginative.
70. Has considerable anxiety over unacceptable sexual impulses.
71. Seems unable to concentrate.
72. Is a friendly and outgoing person.
73. Has a strong need for nurturance and passivity.
74. His hostile impulses are markedly inhibited by his fear of rejection.
75. Is a direct and resourceful person.
76. Has relatively poor intellectual control.
77. His homosexual trends are not a source of great conflict.
78. Feels pretty much on a par with his peer group.
79. Is an asocial person.
80. Has marked conflict between his heterosexual urges and his fear of women.
81. Has a strong heterosexual drive.
82. Is attracted to, yet afraid of his mother.
83. Is markedly depressed by adverse criticism.
84. Is eager for social approval.
85. Feels guilty over his hostile impulses.
86. Sees women as superior and threatening.
87. Shows many objective signs of anxiety.
88. Is a very quiet person.
89. Perceives masculine sexuality as involving hostile behavior.
90. Has a strong fear of aggression.
91. Does not seem to be particularly afraid of anything.
92. Hostility directed toward him does not seem to disturb him very much.
93. Has strong dependency needs.
94. Intellectually he is quite resourceful.
95. Always seems to distrust his own ability.
96. Shows agitated depression.
97. Is easily aroused emotionally.
98. Shows "resigned depression."
99. His sexual needs are intense.
100. Resents the restrictions and dominance of others.
101. Is impulsive.
102. Seldom feels guilty.
103. Sees himself as a rather ineffective person.
104. Admires his father.
105. Indulges in an unusual amount of fantasy.
106. Masturbation is not a problem for him at present.
107. Has an even temperament.
108. Is a fairly optimistic person.
109. Uses fantasy as his major outlet for hostile feelings.
110. Is an energetic person.
111. His range of expression of feelings is fairly wide.
112. In interpersonal conflict, he blames himself very readily.
113. Has strong feelings of guilt over his sexual impulses.
114. Considers women ineffectual and helpless.
115. Always has excuses for his behavior.
116. In interpersonal conflict, he tends to blame the other person.
117. Is conscientious and painstaking.
118. Justifies his feelings by attributing them to others also.

Factual True–False Items

121. His first 3 years were relatively untraumatic.
122. Was raised in foster home or orphanage.
123. Was an only child.
124. Was youngest child in the family.
125. Was oldest child in his family.
126. Has had severe blow on the head.
127. Had a sheltered childhood.
128. Had childhood hobbies.
129. As a child had some limiting disease or disorder.
130. As a child his family moved often.
131. As a child was closer to his father than to mother.
132. His father was a fairly stable, unexcitable person.
133. His mother was a fairly stable, unexcitable person.
134. His father was strict.
135. His mother was strict.
136. Was overprotected by his father.
137. Was overprotected by his mother.
138. Was rejected by his father.
139. Was rejected by his mother.
140. Had an ineffectual father.
141. Had an ineffectual mother.
142. As a child his home was maternally dominated.
143. Relied on his father for guidance.
144. Relied on his mother for guidance.
145. Was given intense corporal punishment by a parent.
146. Felt he did not have the love of one or both parents.
147. Had little contact with his father or mother.
148. His father spent little time with him.
149. His father was not regularly in the home.
150. His father was unemployed for long periods.
151. His father was alcoholic.
152. His mother prejudiced him against his father.
153. As a child felt free to communicate in his home.
154. As a child had enuresis, temper tantrums, nail biting, etc.
155. As a child was rejected by his peers.
156. Had a sibling or friend who was close to him.
157. As a child had an unfavorable social environment.
158. Has been on his own from a fairly early age.
159. As a child was hyperactive.
160. As a child was shy and withdrawn.
161. Ran away from home.
162. As a child had difficulties with other children.
163. As a child got into fights.
164. As a child was member of Boy Scouts, clubs, etc.
165. Had difficulties in school.
166. Made generally poor marks in school.
167. Was generally disinterested in school.
168. Had reading difficulties in school.
169. Was truant from school.
170. Dated girls when he was in high school.
171. Through high school was in relatively few social activities.
172. Participated on school teams.
173. Skipped grades in school.
174. Was given sex information by one of his parents.
175. Had early heterosexual play with peer.
176. Suffered early seduction or early sex trauma.
177. Had first sexual intercourse before puberty.
178. Was the aggressor in his first sexual relationship.
179. Had no sexual relations until after he was 21.
180. Has had intercourse with many women.
181. Has had intercourse with prostitutes.
182. Has suffered from guilt over masturbation.
183. Is married.
184. Has been married more than once.
185. Is living with his parents at present.
186. Is fairly religious.
187. Is excessively religious.
188. Is confused in his religious beliefs.
189. Drinks quite a bit.
190. Has strong ethnic prejudices.
191. Gambles quite a bit.
192. Is a non-conformist in much of his behavior.
193. Is extremely conforming in much of his behavior.
194. Is an overt homosexual.
195. Has had homosexual experiences.
196. Has anxiety attacks.
197. Is easily upset or easily tired.
198. Has nightmares.
199. Has fears or phobias.
200. Has obsessive thoughts or compulsions.
201. Has psychosomatic symptoms.
202. Is very seclusive.
203. Is very suspicious.
204. Hallucinates.
205. Is delusional.
206. Is depressed.
207. Has been in trouble with the law.
208. As an adult has been in fights or brawls.
209. As an adult has no interests or hobbies.
210. Has a definitely established gainful occupation.
211. Worked with father or in father's line of work.
212. Had had several jobs in last few years.
213. Had had to stop work because of psychological symptoms.
214. Was in the Service.
215. Happiest time of his life was while in the Service.
216. Suffered severe combat stress during Service.
217. Was court-martialled while in Service.
218. Had several promotions in the Service.
219. Volunteered for dangerous missions in the Service.
220. Suffered nervous breakdown under Service conditions.

Q-Sort Items

1. He likes and seeks the companionship of women.
2. He usually withdraws in social situations.
3. He is obviously uncomfortable in his contacts with authority.
4. He gives in immediately when attacked.
5. He seems to adapt to his environment quite well.
6. Nearly all of his interpersonal relationships are superficial.
7. He meets aggression from others with counter aggression.
8. He has very few friends.
9. He blames himself when criticized.
10. He has difficulty in relating to people.
11. His contacts with authority are marked by strife.
12. He draws into himself in interpersonal situations.
13. He has quite a few relatively warm interpersonal relationships.
14. He does not invest affect in his interpersonal relationships.
15. He is constantly fighting his environment.
16. He is rather cold and impersonal in his interactions with women.
17. He goes out of his way to please people.
18. He accepts rather passively any impositions made upon him.
19. His social adjustment seems satisfactory.
20. He tends to retreat from the stress and strain of everyday living.
21. He meets the demands of his environment in a rather reasonable fashion.
22. He escapes into fantasy when threatened.
23. He tries to manipulate those around him to serve his own ends.
24. He is uneasy and uncomfortable around other men.
25. He becomes quite aggressive in most social situations.
26. He generally feels accepted by others.
27. He behaves rather passively in most situations.
28. He responds well to most of his environmental demands.
29. He reacts in a hostile and aggressive fashion in his interpersonal contacts.
30. He is unable to recognize affection when it is offered to him.
31. He retreats when people become aggressive toward him.
32. He competes very strongly with father figures.
33. He gets along quite well with his superiors.
34. He withdraws immediately at the slightest hint of rejection.
35. He becomes actively hostile when he is criticized.
36. He is only mildly responsive to friendly approaches from others.
37. He is generally composed and at ease with authority figures.
38. He is quite responsive when people show an interest in his welfare.
39. He faces unpleasant situations in a direct manner.
40. His immediate response when threatened is to fight back.
41. He gets along well with most people.
42. He responds gratefully to positive gestures from others.
43. He is quite passive in most of his interpersonal contacts.
44. He adapts easily to social situations.
45. His friendships seem rather stable and have relatively few conflicts.
46. His relations with mother figures are fairly happy.
47. He yields meekly to the demands of others.
48. He responds immediately to any show of tenderness or affection.
49. Rejection from others only makes him more aggressive.
50. He is tactful and considerate with other people.
51. He runs away from hostility in others.
52. He seems to handle criticism quite well.
53. Friendly gestures from others arouse his suspicion and hostility.
54. He is quite responsive to social stimuli.
55. He rarely responds to positive gestures from others.
56. He runs away from any show of tenderness or affection.
57. When others show an interest in him he immediately becomes aggressive and demanding.
58. He has little difficulty in dealing with other people.
59. He is responsive to the needs and wishes of others.
60. He likes to have people show an interest in him.
61. He avoids any display of friendliness on the part of others.
62. He responds in a realistic fashion to hostility from others.
63. He responds passively to any positive gestures made toward him.
64. He actively seeks social contacts.
65. Even the mere presence of other people produces immediate withdrawal in him.
66. He tries to dominate other people.
67. He gets along quite well with other men.
68. He actively rejects affection from others.
69. Almost any social situation arouses aggressive behavior on his part.
70. He seems quite meek in most of his interactions with others.
71. He shows decreasing involvement in social contacts.
72. He adapts fairly easily to authority demands.
73. He becomes negativistic when people show a friendly interest in him.
74. He is rather neutral towards others.
75. He is rather reserved in most situations.
76. He accepts affection passively.

APPENDIX D

CATEGORIES OF REPORT-TYPE ITEMS

1. Pressures; forces; press
2. Motivations; goals; drives
3. Outlooks; attitudes; beliefs
4. Frustrations; conflicts; fears
5. Affects; feelings; emotions
6. Sexual thought and behavior
7. Psychosexual level and development
8. Super-ego; values; ego ideals
9. Self control; ego strength; ego capacity
10. Self concept; insight into self
11. Personality defenses and personality mechanisms
12. Reality contact; orientation
13. Interpersonal and object relations
14. Quality of perception, fantasy, language, and thought
15. Intellect and abilities
16. Symptoms; diagnoses; etiology
17. Prognoses; predictions
18. Postdictions

APPENDIX E

SIMILARITY RATING SCALE

INSTRUCTIONS

On the following pages are a number of *pairs* of diagnostic statements. Your task is to judge the *similarity* of the two statements within each pair and to indicate your judgment by writing *one* of the six numbers following each pair of statements. In general, your judgment should reflect the extent to which you would consider both diagnoses reasonable and acceptable for one patient when made by two equally competent diagnosticians from the same thorough examination.

The scale is defined as follows:

5—Identical; perhaps minor differences in wording but no question but that the two statements in the pair mean the same thing.

4—Markedly similar; slight variations in minor symptomatology which would ordinarily be considered unimportant.

3—Similar; same general category but definite differences in implied or stated significant characteristics.

2—Different; diagnoses from different but adjoining broad categories; would require resolution of the discrepancies.

1—Markedly different; very rarely would one expect to see two such diagnoses on the same person; both cannot be accepted.

0—Utterly different; improbable that two competent diagnosticians could arrive at such opposite conclusions.

You will notice that some diagnoses are repeated but are paired with different diagnoses. Occasionally even a pair may be repeated. In such cases, you may feel free to refer back to your previous judgment. Do not linger too long on any one pair; it is your general clinical judgment that is desired.

Vol. 73, No. 8 Whole No. 478, 1959

Psychological Monographs: General and Applied

THE INFLUENCE OF CLASSIFICATION REQUIREMENTS ON GRADIENTS OF RESPONSE[1]

MICHAEL A. WALLACH

Harvard University

WE BEGIN with the observation that a person is constantly classifying the contents of the world about him. The child, for instance, is taught what kinds of stimulation may be called green and what ones blue. Later he may likewise learn what modern Italian paintings are futurist in style rather than surrealist, or what people should be called communists. There seem to be certain common features involved in all such classifying, and they may be put as follows. Stimulation is received which the individual arrays along one or more innate or learned dimensions. Such input, along wider or narrower extents of such dimensions, is associated by repeated contiguity with a class name. It is in such a manner as this that the individual comes to learn what parts of the world of stimulation it is appropriate to call of yellow color, politically anarchistic, or impressionistic in style of painting.

But the learning situation is often rife with elements of ambiguity. Was that color blue or green? Its hue was such that one cannot be sure. Here is a painting whose style seems to reflect both futurist and surrealist components. Are this man's activities really of a sort that justifies calling him a communist? With our world of fleeting stimuli, of quick looks, and conflicting opinions as to what class labels are appropriate in what situations, such ambiguities as these would seem to be common indeed. If one is to study how people learn to classify, then, it would seem necessary from the outset to recognize that the world more often than not provides a person with stimulation that does not clearly fit the definition of any one class or label, but rather stands at the boundary of a particular label's applicability; it can be labeled with equal or near equal appropriateness by more than one term. Thus we come to the central question of our inquiry: Do any factors arising from the nature of classification itself influence how it proceeds in an ambiguous situation—i.e., a situation where alternative class names for the given event have been learned to equal degree?

There are at least two mechanisms arising from the way classes are defined which influence the tendency to assign information to a particular class: (*a*) whether a class is defined by inclusion or exclusion and (*b*) whether a class defined by inclusion is contrasted with a class defined by exclusion of the former class or with one also defined by inclusion. Consider the first question. There seem to be two main ways in which one can specify the contents of a class, and our hypothesis is that their psychological implications are quite different. On the one hand, the contents of a class may be specified by inclusion; such-and-such attribute values must be present for the datum to be given a particular class label. If, for instance, something has wings, two legs, and flies through the air, we may appropriately

---

[1] This monograph is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Harvard University. The author is greatly indebted to Jerome S. Bruner, his principal adviser, for encouragement and aid in this research. He also wishes to thank Gordon W. Allport, John B. Carroll, George Mandler, and Evon Z. Vogt, the other members of his Thesis Committee, and Albert J. Caron, Paul Friedrich, and Frederick Mosteller, all of whom rendered assistance in various ways. Special thanks are due Rainer Weiss for design of the electronic apparatus. This research was conducted while the author held a doctoral research fellowship from the Samuel S. Fels Foundation. Support was also provided by Grant M-1324 from the United States Public Health Service, under the general direction of Jerome S. Bruner.

No. 7 wanting

classify it as a bird. On the other hand, the contents of a class may be specified by exclusion; if such-and-such attribute values are not present, the datum can be called a nonexemplar of the class or classes appropriate to those attribute values. If, for example, an animal does not spend all his time on land and/or in the water, does not have more than two legs, does not lack wings, then we may appropriately classify it as an animal residing somewhere other than on land or sea—and we will be referring here to the same domain of instances that we labeled before as "birds."

Given an inclusion and an exclusion class that both refer to the same domain of instances, our hypothesis is that the range of acceptance limits for placing data in a class will be greater for the former than for the latter. An inclusion class, we suggest, carries with it the implication that it is well represented in the environment—that there are many events for which the class label in question has clear applicability. An exclusion class, on the other hand, is a more vague entity. It is as if people prefer examples of what something is, more than examples of what something is not. This in turn might be expected on the ground, for instance, that it is usually difficult to know how to deal with an event placed in an exclusion class since the event could be any of a variety of things, or that work with an event defined by exclusion usually requires the added cognitive effort of transformation to a definition by inclusion.

Our second suggested classification determinant of range of acceptance limits is whether an inclusion-defined class of interest is being contrasted with a class also defined by inclusion or with the contrast class defined by exclusion of that class of interest. We would propose that reference to some contrast class is always present, implicitly if not explicitly, when we classify, and this for the following reason. In order to avoid being overwhelmed by the confusing richness of attributes that the environment offers the senses, a person must define a class for himself in terms of a limited selection of properties. Consideration of a contrast class, in turn, usually provides one mechanism for such selection. If we make setters the contrast when defining the class "cocker spaniels," for instance, then we may tend to take size as criterial for our definition of the latter class. If, however, we take "all animals other than cockers" as the implied contrast, then we may well define cockers differently. Thus a contrast class too may be defined by inclusion or exclusion, and we hypothesize that this difference is a second determinant of the range of acceptance limits for placing events in an inclusion class of interest. If the contrast class is defined by exclusion of the class of interest, we predict that the acceptance limits for placing events in that inclusion class of interest will be broader than if the contrast class is another inclusion class. This should follow because we expect a contrast class defined by exclusion of a class to be relied on to a lesser extent than an inclusion-defined contrast class, hence giving the class of interest more opportunity in the former case to capture events away from the contrast class.

VARIABLES AND HYPOTHESES UNDER STUDY

Dependent Variable

Range of acceptance limits for placing events in a class. The range of acceptance limits for placing events in a class is said to be greater, the greater the number of discriminably different events assigned to that class, and the higher the relative frequency of assignment of each such event to one class rather than to another class or other classes (or the greater the magnitude of some response indicating assignment to a class).

Independent Variables

Definition of a class of interest by inclusion or exclusion (positive or negative class). A domain of data is said to be defined by inclusion, or assigned to a positive class, if the definition of the class specifies the particular attribute values whose presence is definitive of instances of that class. A domain of data is said to be defined by exclusion, or assigned to a negative class, if the definition of the class specifies simply that the attribute values definitive of instances of some other class or classes are absent—hence, that instances of some positive class or classes are absent.

Presence of a positive or complementary negative contrast class for a positive class of interest. The class considered as a contrast to a positive class of interest may be the negative class complementary to that positive class of interest (if the positive class of interest is called "X," the complementary negative contrast class is "$\bar{X}$"), or the contrast class may be some other positively defined class.

Hypothesized Relationships Between Independent Variables and the Dependent Variable

Hypothesis 1: Concerning the influence of positive vs. negative classes on the range of acceptance limits. (*a*) For a class whose definition is never more than ambiguous, a positive class has a greater range of acceptance limits for placing events in it than does a negative class associated with the same domain of events. (*b*) The same is true for a class whose definition is at least in part certain.

Hypothesis 2: Concerning the influence of a contrast class, and how it is defined, on the range of acceptance limits for the class of interest. (*a*) For a positive class whose definition is never more than ambiguous, the range of acceptance limits for placing events in that class will be greater when contrasted with its complementary negative class than when contrasted with another positive class. (*b*) The same is true for a positive class of interest whose definition is at least in part certain.

It should be noted that we expect these relationships only to hold with regard to data that are generally ambiguous. Data are maximally ambiguous if their relative frequencies of assignment to alternative classes are the same. Data are maximally certain, on the other hand, if their relative frequency of assignment to one class is 100%. We are interested in the determinants of *S*'s classification of generally ambiguous data because such ambiguity is the case in those life situations for which the present hypotheses are supposed to be paradigmatic.

Now a class may be ambiguously associated with some data and associated more certainly with other data, or it may be only ambiguously associated with data. The question thus arises of whether the effects we postulate will hold only with regard to a class which is itself never more than ambiguous, or rather will also hold regarding a class whose definition is at least in part certain. Hence Parts *a* and *b* of each hypothesis. A class is maximally ambiguous, then, if no datum has more than an equal relative frequency of assignment to that class and to one or more other classes, and the greater the number of discriminable data of this kind. On the other hand, a class becomes increasingly certain, the greater the relative frequency of assignment of data to that class rather than to another class or classes, and the greater the number of discriminable data with a high relative frequency of assignment of this kind.

Method

Design of Conditions

We turn now to the matter of testing the hypotheses presented. The experimental requirements are these: (*a*) We must vary the way a class of interest is defined—positively vs. negatively—meanwhile holding the contrast class constant. The logical impossibility of a class of interest and a contrast class both being negative may be noted.[2] Hence, it follows that the contrast class held constant here must be positive, while the class of interest varies from positive to negative. (*b*) We must also vary the way the contrast class is defined, meanwhile holding the type of definition of the class of interest constant. Since we wish to compare the case where the contrast class is positively defined with the case where it is defined by exclusion of the class of interest, it follows that the class of interest held constant must be defined positively—again to avoid the logically impossible case of contrast class and class of interest both being negative. Further, we want situations where the classes whose acceptance limits we are determining are never more than ambiguous in definition, and other situations where these classes' definitions are in part certain.

---

[2] It would be logically impossible to have both class of interest and contrast class negative, because two negative classes cannot apply to a domain when the two classes are mutually exclusive (i.e., cannot both apply to the same event at the same time) and are all-inclusive.

Let us refer to a class defined positively as "X" or "Y," and to one defined negatively as "$\bar{X}$" or "$\bar{Y}$." Three conditions, involving association of these classes with a range of data in different ways, will permit us to test the above hypotheses. These conditions are as follows:

The XY condition. Both classes associated with the domain of data are positively defined. The association of Class X with that domain is in part certain, while the association of Class Y with that domain is never more than ambiguous.

The $X\bar{X}$ condition. A positive class and its complementary negative class are associated with the domain of data. The association of the positive class (X) with that domain is in part certain, while the association of the negative class ($\bar{X}$) with that domain is never more than ambiguous.

The $\bar{Y}Y$ condition. Again, a positive class and its complementary negative class are associated with the data domain. This time, however, the association of the positive class (Y) with that domain is never more than ambiguous, while the association of its complementary negative ($\bar{Y}$) is in part certain.

How do these conditions permit us to test our hypotheses? Consider Hypothesis 1*a*. If we compare Conditions XY and $X\bar{X}$, we are holding a contrast class constant and positive (X), and varying the definition of an ambiguous class of interest from positive (Y) to negative ($\bar{X}$)—which is what must be done to test that hypothesis. For Hypothesis 1*b*, if we compare conditions XY and $\bar{Y}Y$, we are holding a contrast class constant and positive (Y), and are varying the definition of a class of interest which is in part certain from positive (X) to negative ($\bar{Y}$)—the requirement for testing that hypothesis. For Hypothesis 2*a*, we again compare Conditions XY and $\bar{Y}Y$, since here the ambiguous positive class of interest is Y, and we compare a positively defined contrast class (X) and a complementary negative contrast class ($\bar{Y}$). Finally, in the case of Hypothesis 2*b*, we consider Conditions XY and $X\bar{X}$ again, for here the positive and, in part, certain class of interest is X, and we compare a positively defined contrast class (Y) and a complementary negative contrast class ($\bar{X}$).

Note that the groups being compared are the same for Hypotheses 1*a* and 2*b*, and for Hypotheses 1*b* and 2*a*. How, then, can each hypothesis in the pair be tested separately if the same conditions are compared for the two hypotheses? The answer lies in consideration of the measures we use to determine range of acceptance limits. One kind of measure involves *S*'s predicting, for each successively presented stimulus event, whether it will be given one or the other class label, or his refusing to predict. Another kind of measure, after *S* has experienced the whole set of associations between various stimulus events and the two class labels, involves *S*'s reporting, for each successively presented datum, whether it was given one or the other class label more than half the time before, or was given each label about half the time before. Provision of the "refusal to predict" response choice in the first kind of measure, and of the "half and half" response choice in the second kind of measure, permit separate testing of each hypothesis in the pairs just noted, even though both in the pair are tested by comparing the same two conditions. This comes about because these response choices allow the relative frequency of assignment of data to Class X to vary independently of the relative frequency of assignment of data to Class Y for the XY condition; likewise, the relative frequencies of data assignment to Classes X and $\bar{X}$ can vary independently in the $X\bar{X}$ condition, and those to $\bar{Y}$ and Y can vary independently in the $\bar{Y}Y$ condition. These frequencies are independent since there is always a third choice which *S* can make in any of the groups: he can refuse to offer a prediction or can say each label occurred about half the time before. (Let us abbreviate responses of the latter two sorts as "O" judgments.) Calling the two class labels in each case (i.e., X vs. Y, X vs. $\bar{X}$, $\bar{Y}$ vs. Y) Labels 1 and 2, respectively, we hence see that assignment of input to Label 1 can change in relative frequency without a reciprocal change in relative frequency of assignment of input to Label 2 (or vice versa), because there is a "cushion" in between: namely, "O" judgments.

Consider now what each of the hypotheses predicts concerning these relative frequencies. According to Hypothesis 1*a*, comparing XY and $X\bar{X}$, the relative frequency of assignment of data to Class Y should be greater than that to Class $\bar{X}$. By Hypothesis 1*b*, comparing XY and $\bar{Y}Y$ conditions, the relative frequency of assignment of data to Class X should be greater than that to Class $\bar{Y}$. By Hypothesis 2*a*, again comparing XY and $\bar{Y}Y$ conditions, the relative frequency of assignment of data to Class Y should be greater for the $\bar{Y}Y$ group, and less for the XY group. Finally, by Hypothesis 2*b*, again comparing the XY and $X\bar{X}$ groups, the relative frequency of assignment of data to Class X should be greater in the $X\bar{X}$ group, and less in the XY group. It becomes clear, then, that the four hypotheses can be independently tested if, as is the case, the relative frequencies of data assignment to Classes X vs. Y, X vs. $\bar{X}$, and $\bar{Y}$ vs. Y, are independent.

Experimental Procedure

Light intensity was used as the dimension of stimulus input, since it was a simple one and changes in its values could be easily created and measured. The *S*'s training regimen involved associating brightnesses with labels or class names.

The questions of defining a class positively vs. negatively, and providing a contrast class defined positively or defined by exclusion of the positive class of interest, were handled by varying the kinds of class labels provided. For Condition XY, where both labels were to be positive, when a brightness was to be linked with Label 1, a bulb lit up the letter "X"; and when a brightness was to be linked with Label 2, a different bulb lit up the letter "Y"—the "X" and "Y" bulbs being placed one on each side of the screen where brightnesses were presented. When there are separate "X" and "Y" bulbs and one of them lights each time, we thereby imply to *S* that there are two classes each defined by particular values of the brightness attribute—i.e., two positive classes.

For Conditions $X\bar{X}$ and $\bar{Y}Y$ where, on the other hand, one label was to be positive and the other its complementary negative, when a brightness was to be linked with the positive label, a bulb lit; while when a brightness was to be linked with the negative label, that same bulb failed to light—the single bulb being placed underneath the screen where brightnesses were presented, and no other bulb being present. When there is just one bulb which either lights or fails to light each time, we thereby imply to *S* that there is one class defined by particular brightness values and another class defined simply by absence of those values—i.e., a positive class and a complementary negative class, respectively.

Note that the actual stimulus inputs of brightnesses and the relative frequencies of particular associations of brightnesses with one of two possible class labels, were alike in all conditions. What varied were the instructions provided by the light labels as to how the classes were defined, while the structures of the three pairs of classes were exactly the same.

The number of different brightness values used was 16, these being spaced on a scale of just noticeable differences (see Table 1).[3] The total number of presentations of brightnesses during the training series was 160, each brightness coming up 10 times and the whole sequence being uniquely random, but every brightness occurring twice in each block of 32 presentations. The relative fre-

---

[3] The JND's were established by trials of the following sort: A rheostat was slowly turned upward from minimal brightness (or downward from maximal brightness), brightness intensity level being recorded every time *S* reported a just noticeable difference. Two upward series and two downward series were given in single alternation to any one *S*. Sixteen was about the maximal number of JND's detectable within the brightness range available from the rheostat. The *E* ran this procedure on himself and also on two college student *S*s. Final brightness intensity settings were arrived at by taking all series in which 16 brightness values had been provided by the judge (i.e., 15 JND's) and finding the median of the successive intensity settings for each of the 16 values. Sixteen brightness values were provided in three of the four series by one judge, and in two of the four series by each of the two other judges. (Fourteen or 15 different values were provided in the remaining series.)

TABLE 1

Brightness Levels Used

Brightness Intensity (In DC ma. using a 60-w. bulb)	Brightness No.
1.00 (Brightest)	1
.98	2
.96	3
.88	4
.84	5
.80	6
.76	7
.72	8
.68	9
.63	10
.57	11
.53	12
.48	13
.43	14
.40	15
.36 (Darkest)	16

quencies of association of each brightness with each of the two labels used in any one condition are shown in Fig. 1. As indicated on that figure, toward the brighter end of the brightness dimension all presentations were always associated with one label, while toward the darker end, all presentations were equally often associated with each of the two labels. The sequence of occurrence of the two labels in any condition was uniquely random, and was the same for the label of the same number (1 or 2) in all the conditions.

Seventy persons participated in the study, the number of university student *S*s in Groups XY and XX̄ being 23 each, while Group ȲY contained 24 *S*s. Procedure was the same for all three groups except for the labeling differences among these groups as discussed above. This procedure fell into three parts and was arrived at as a result of pilot investigation.

Training series. On each of the 160 trials, one of the 16 brightnesses was flashed for 0.5 sec. in a uniquely random order. After each brightness was presented, *S* pushed one of three buttons, indicating, if in Group XY, that he bets the "X" light will go on, the "Y" light will go on, or that he refrains from making a bet, respectively; if in Groups XX̄ or ȲY, that he bets the single light will go on ("Yes"), the single light won't go on ("No"), or that he refrains from making a bet, respectively. When *S* pushed any button, the "X" or "Y" light went on for 1.0 sec. if in Group XY, or the single light went on for 1.0 sec. or remained off if in Groups XX̄ or ȲY—which of the two class labels occurred depending on the brightness that occurred on that trial. After a pause of 5.5 sec. following a button push, the next brightness in the sequence flashed, thereby beginning the next trial. Thus, in the course of the 160 training trials, each brightness was linked with one or both labels according to the pattern of association diagrammed in Fig. 1.

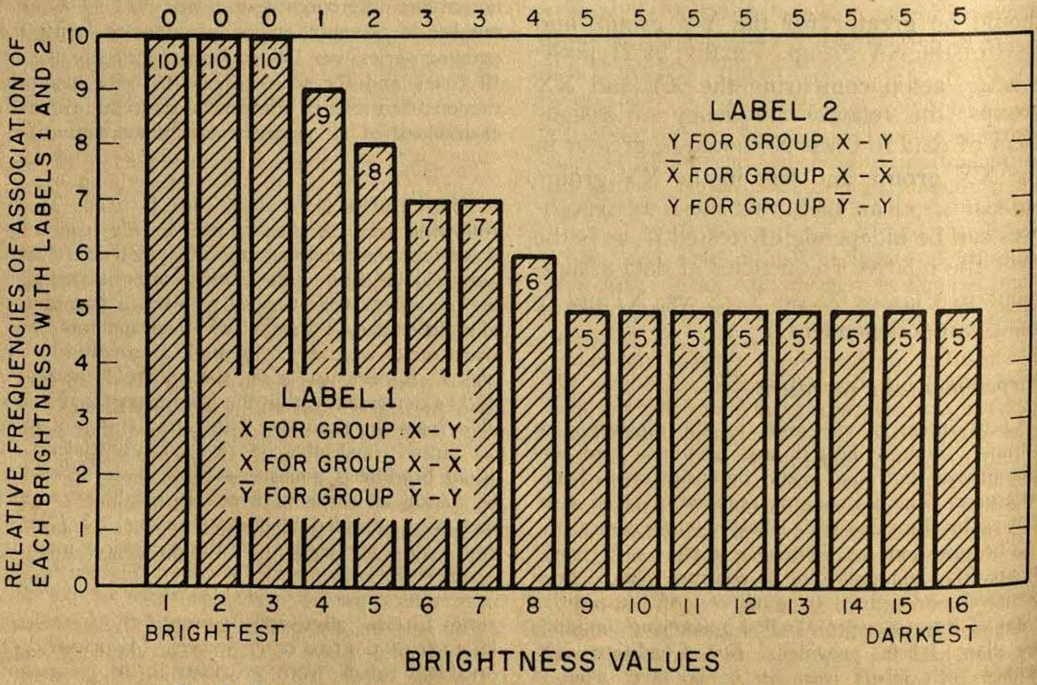


Fig. 1. Degrees of association of each brightness with Labels 1 and 2.

Test with brightnesses. On each of 32 trials, one of the 16 brightnesses was flashed for 0.5 sec. in a uniquely random order, each brightness occurring twice during the series. After each brightness was presented, *S*, if in Group XY, indicated, by pushing a button labeled "More," that he thinks this brightness turned out to be X more than half of the times it was previously presented or turned out to be Y more than half those times, and then told *E* verbally whether he meant X or Y more than half the time before; or *S* might push another button labeled "Equal," to indicate that he thinks this brightness turned out to be X about half those times and Y about half those times. If in Groups $X\bar{X}$ or $\bar{Y}Y$, *S* indicated, by pushing the button labeled "More," that he thinks this brightness caused the light to go on more than half of the times it was previously presented or caused the light to stay off more than half of those times, and then told *E* verbally which he meant; or *S* might push the other button labeled "Equal" to indicate that he feels this brightness caused the light to go on about half those times and to stay off the other half. No label lights went on during this test series. Five and one-half seconds after *S* pushed a button, the next brightness in the sequence flashed, thus starting the next trial.[4]

Test without brightnesses. The *S* was shown a sheet of paper along which a horizontal line was drawn to represent the dimension of brightnesses to which he had been exposed, the word "Darkest" appearing at one end of the line and "Brightest" at the other.[5] Along a vertical line at the left, divided into 10% steps ranging from 0% to 100%, appeared the legend "Proportion of times you think each brightness was labeled 'X'" if for Group XY; while the legend for Groups $X\bar{X}$ and $\bar{Y}Y$ read "Proportion of times you think each brightness made the light go on." The *S* was requested to draw a gradient indicating the proportion of times he thinks each brightness was labeled "X" (Group XY), or made the light go on (Groups $X\bar{X}$ and $\bar{Y}Y$), for all parts of the brightness dimension.

The choice measure in the training series indicated the learning of labeling tendencies. The choice measure in the test with brightnesses provided a more perceptually based indication of the labeling tendencies thus learned (since the brightnesses were presented again), while the graph measure in the test without brightnesses provided a more conceptually based index of the same thing (since the brightnesses were not presented). Latency measures in the training series and the test with brightnesses told us how confident *S* was in his labeling decisions.

Apparatus

The time lengths for presentation of the brightnesses and of the label lights, and for the interval between *S*'s pushing of a button and the occurrence of the next brightness in the sequence, were automatically controlled by timers. Pre-established programs on stepping switches automatically controlled the sequences in which the brightnesses occurred and in which the two labels in any condition occurred. The brightness differences were differences in the voltage settings of 16 rheostats, some one of which controlled the lighting of a common incandescent bulb for each trial. This high degree of automatic control of the sequence and timing of stimulus and label materials was required in order to free *E* to record the various items of data that *S* provided, while still maintaining strict temporal control over presentation of stimulus input and label materials.

The sequence of events in time is shown in Fig. 2, and the block diagram of Fig. 3 indicates the various apparatus components and how they were linked so as to result in this sequence. The pushbuttons for the test series were mounted on one portable board while those for the training series were mounted on another, and one set of pushbuttons was placed out of *S*'s sight when he was

---

[4] Use of the two buttons noted in this test without brightnesses, rather than providing three buttons for each of the three alternative decisions, was instituted as an added control to insure that *S* was pushed toward veridicality of report rather than away from it. A single button labeled "More" and another one labeled "Equal" imply to *S* that there might well be some brightnesses linked to one label more than half the time, and about as many other brightnesses linked to each label equal proportions of the time—which is in fact the case. The implication tendered, in other words, is that there is warrant to press the "Equal" button on about half of the test trials. To follow this implication would lead *S* toward veridicality. If nonveridicality is obtained, then, in that Label 2 is thought to apply more often than was the case, this tendency was not fostered by the mere physical conditions of the test itself.

The instructions given the various groups again insured that *S* was pushed toward veridicality of report rather than away from it, in that the examples given of kinds of label-input associations told only of Label 1 being linked with a brightness shade all the times that shade is presented, or most of the times it is presented, or about half the times. No example was mentioned where Label 2 was said to be linked with a brightness shade more than half the time, although of course this possibility was not ruled out. Once again, if nonveridicality is obtained in that Label 2 is thought to apply more than half the time to some brightness, then this decision was not fostered by bias in the communicated instructions.

[5] For half the *S*s in each group, "Darkest" appeared at the left and "Brightest" at the right end of the line, while for the other half these word positions were reversed.

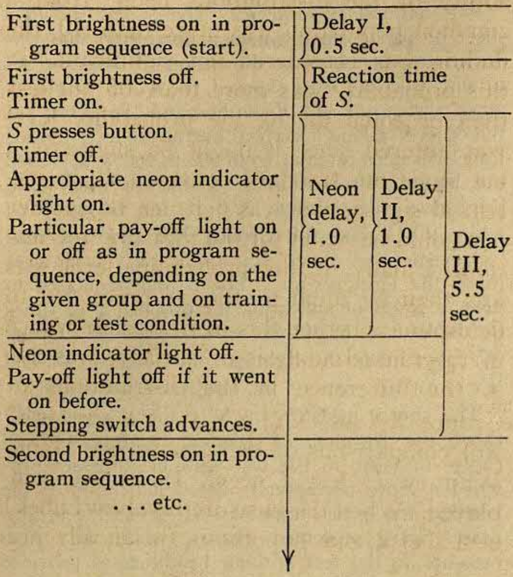

First brightness on in program sequence (start).
Delay I, 0.5 sec.
First brightness off.
Timer on.
Reaction time of S.
S presses button.
Timer off.
Appropriate neon indicator light on.
Particular pay-off light on or off as in program sequence, depending on the given group and on training or test condition.
Neon delay, 1.0 sec.
Delay II, 1.0 sec.
Delay III, 5.5 sec.
Neon indicator light off.
Pay-off light off if it went on before.
Stepping switch advances.
Second brightness on in program sequence.
. . . etc.

FIG. 2. Sequence of events in time.

using the other. Neon indicator lights were mounted on *E*'s panel to tell him which button *S* pushed on that trial. The neon delays resulted in a neon light's staying on for 1 sec. after *S* released the button, so *E* was sure to see which light went on. The *E* began the brightness sequence in the training or test series by pressing the start button, and from then on the series proceeded automatically. When the end of the series was reached and *E* wished to set up the sequence for the next part of the experiment or for the next *S*, *E* returned the program sequences to their starting position by pressing the reset button, which homes all the stepping switches. The timer utilized was a Standard Electric Time Co. product, Type S-1. A regular 60-w. lamp of white color was the brightness intensity bulb, and it was mounted behind a circle of translucent blue celluloid 1.5 in. in diameter, so brightness differences were various intensities of blue light. Pay-off lights were 3-w. lamps mounted in pilot light fixtures.

The experiment was run in semidarkness, room lights being off and the only source of illumination other than the brightness intensities and pay-off lights being a small 25-w. reading lamp placed well behind a black curtain on *E*'s desk.[6] The *S* sat 3 ft. from the light panel and at eye level with it.

Figure 4 shows a sketch of the apparatus from *S*'s point of view.

RESULTS

Tables 2–6 summarize the results obtained using the measures outlined above. It will be noticed that when results are presented concerning various parts of the whole

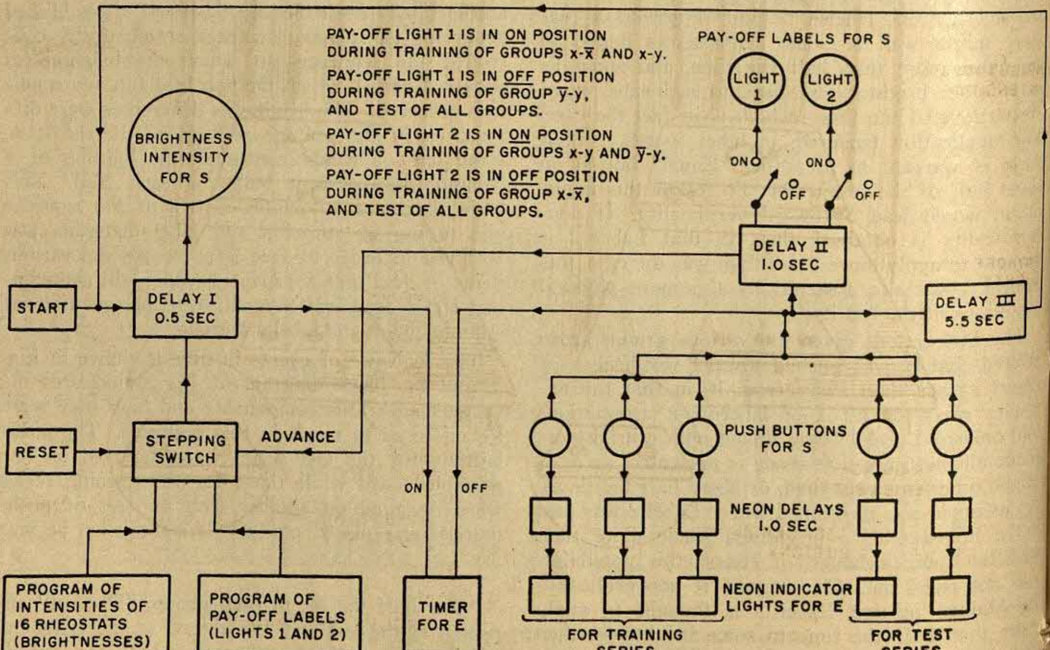

PAY-OFF LIGHT 1 IS IN ON POSITION DURING TRAINING OF GROUPS x-x̄ AND x-y.
PAY-OFF LIGHT 1 IS IN OFF POSITION DURING TRAINING OF GROUP ȳ-y, AND TEST OF ALL GROUPS.
PAY-OFF LIGHT 2 IS IN ON POSITION DURING TRAINING OF GROUPS x-y AND ȳ-y.
PAY-OFF LIGHT 2 IS IN OFF POSITION DURING TRAINING OF GROUP x-x̄, AND TEST OF ALL GROUPS.
PAY-OFF LABELS FOR S
LIGHT 1
LIGHT 2
ON
ON
OFF
OFF
BRIGHTNESS INTENSITY FOR S
DELAY II 1.0 SEC
START
DELAY I 0.5 SEC
DELAY III 5.5 SEC
RESET
STEPPING SWITCH
ADVANCE
ON
OFF
PUSH BUTTONS FOR S
NEON DELAYS 1.0 SEC
NEON INDICATOR LIGHTS FOR E
PROGRAM OF INTENSITIES OF 16 RHEOSTATS (BRIGHTNESSES)
PROGRAM OF PAY-OFF LABELS (LIGHTS 1 AND 2)
TIMER FOR E
FOR TRAINING SERIES
FOR TEST SERIES

FIG. 3. Block diagram of the components of the apparatus.

---

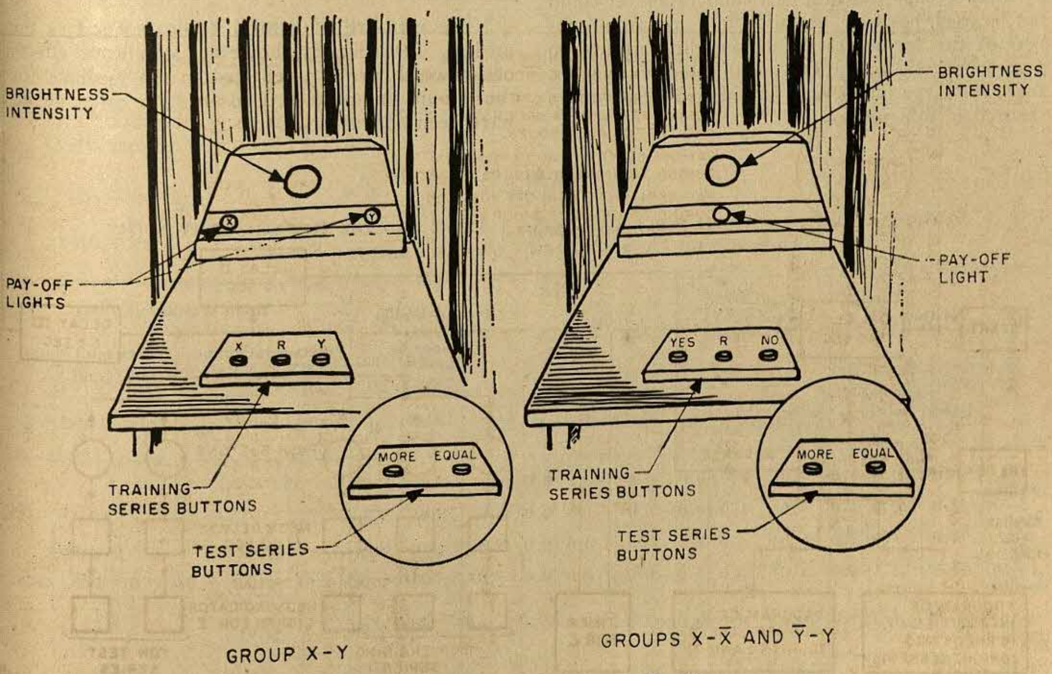
[6] This semidarkness was maintained during the training series, during the instructions for the test with brightnesses, and during this latter test itself.

dimension of 16 brightnesses, we fractionate that dimension into four parts of four brightnesses each and perform calculations for each part separately, as well as examining the effect for all 16 brightnesses taken as a whole. Our reason for breaking the brightness dimension down into no more and no less than four parts, when examining trends across the dimension, is that four was the modal and median number of different brightnesses which our *S*s discriminated in the course of brightness presentations in the two random series. After finishing the test with brightnesses and before starting the test without brightnesses, each *S* was asked how many different brightnesses he thought had been used—and the mode and median reply they gave was four. It hence seemed most psychologically meaningful to take this estimate as the number of segments into which to break the brightness dimension for analysis.

Turning now to the findings in Tables 2–5: For any measure, the results for the three groups of the experiment were first treated by the Kruskal-Wallis ranked analysis of variance (Siegel, 1956). If the probability of the distributions being from a common population was more than .10, the finding was considered not significant. If this probability was more than .05 but less than or equal to .10, the probability level was entered since it might be suggestive; but since the standard convention for statistical significance was not met, no further test could be made for the measure in question. Finally, if the probability level was less than or equal to .05, we proceeded to determine whether this over-all significance of the ranked analysis of variance was due to the difference in the distributions of Groups $X\bar{X}$ vs. XY, XY vs. $\bar{Y}Y$, or both.[7] For comparisons of two groups, the Mann-Whitney *U* test (Siegel, 1956) was employed. When the measure involves Label 1 or Label 2, our hypothesis specifically pre-

---

[7] Such a procedure is analogous to following up the finding that a parametric analysis of variance is significant by using Tukey's test for a significant gap to determine which intergroup differences cause the significant over-all effect. One is, of course, warranted to do tests concerning a pair of the groups only if the over-all effect for the three groups is significant.



GROUP X-Y

GROUPS X-$\bar{X}$ AND $\bar{Y}$-Y

FIG. 4. The apparatus as seen by *S*.

dicts the direction of the difference between Groups $X\bar{X}$ and XY, or between Groups XY and $\bar{Y}Y$. Hence, when a Mann-Whitney test concerns Label 1 or Label 2, a one-tail test was made. When, on the other hand, the measure concerns refraining, "equal" judgments, or the indifference point, we have alternative hypotheses which could predict either direction of a difference, and hence a two-tail test was made.[8]

We offer the reminder that our hypotheses are expected to hold most strongly concerning the more ambiguous half of the brightness dimension—namely, Brightnesses 9–16—since this is the area at the boundary of the category defined by Label 1. Our findings indicate that such, by and large, is the case: The hypotheses noted earlier are usually confirmed, in whole or in part, for Segments 9–12 and 13–16 of the brightness dimension, and are usually not confirmed for Portions 1–4 and 5–8. It hence requires a quite strong effect over Brightnesses 9–16 for our hypotheses to be confirmed when the whole dimension—Brightnesses 1–16—is considered as a unit: yet such confirmation often occurs to at least some extent.[9]

Label Learning (Choice Measure, Training Series)

Tests of the frequencies of choice of Label 1, Label 2, and of the refraining button during the last 32 trials of the training series, are presented in Table 2. From the choice frequencies for Label 1, we note that Hypothesis 1*b*, predicting XY > $\bar{Y}Y$, is supported at the .07 level for Brightnesses 13–16, and Hypothesis 2*b*, predicting $X\bar{X}$ > XY, is supported at the .01 level for those brightnesses. Neither hypothesis is supported for other parts of the brightness dimension. The choice frequencies for Label 2 indicate that Hypothesis 1*a*, predicting $X\bar{X}$ < XY, is supported at the .09 level for Brightnesses 9–12, and at the .001 level for Brightnesses 13–16. Hypothesis 2*a*, predicting XY < $\bar{Y}Y$, is supported at the .06 level for 9–12, but not for 13–16. The over-all effect for Brightnesses 1–16 supports Hypothesis 1*a* at the .01 level, and

---

[8] Whitney's extension of the *U* test to three groups (Mosteller & Bush, 1954) might seem appropriate for Tables 2–5, since this test permits one to evaluate the hypothesis that the distributions of three groups fall in a particular order. It could not be used, however, because there are large numbers of ties for scores among our three groups on a measure, and application of this technique when there are large numbers of intergroup ties involves theoretical difficulties.

[9] The measure of central tendency used is the median. Occasionally, however, significant effects occur which are not reflected in the medians for the groups in question, and in such cases means have been added.

TABLE 2

MEDIAN FREQUENCY OF CHOICE OF LABELS 1 AND 2 AND THE REFRAIN BUTTON (R) DURING THE LAST 32 TRIALS OF THE TRAINING SERIES

Brightness	Group $X\bar{X}$ ($N = 23$)			Group XY ($N = 23$)			Group $\bar{Y}Y$ ($N = 24$)			Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney *U* Tests					
													Between Groups $X\bar{X}$ and XY			Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4, median	8.0	0.0	0.0	7.0	0.0	0.0	6.5	0.0	0.0	N.S.	N.S.	N.S.						
5–8, median	5.0	0.0	1.0	7.0	1.0	0.0	5.0	1.0	0.0	N.S.	N.S.	N.S.						
9–12, median	4.0	1.0	2.0	5.0	2.0	1.0	3.5	3.0	0.0	N.S.	$\leq .05$	N.S.		$\leq .09$			$\leq .06$	
13–16, median	3.0	3.1	2.0	1.0	6.0	1.0	0.0	6.0	0.0	$\leq .01$	$\leq .001$	$\leq .02$	$\leq .01$	$\leq .001$	$\leq .03$	$\leq .07$	N.S.	N.S.
1–16, median	19.0	5.0	7.0	20.0	9.0	2.0	15.5	11.0	2.0	$\leq .08$	$\leq .001$	$\leq .08$		$\leq .01$			$\leq .03$	
1–16, mean	19.7			18.9			16.2											

supports Hypothesis 2*a* at the .03 level. Regarding the frequencies of refraining, we note that $X\bar{X} > XY$ at the .03 level for Brightnesses 13–16, a finding that supports Hypothesis 1*a* and infirms Hypothesis 2*b*. The findings of no significant difference for comparisons of other groups over Brightnesses 9–12 and 13–16 may be supportive, in general, of all hypotheses in question, or may, on the other hand, simply be due to nonconfirmation of these hypotheses. Interpretation of failure to find a significant difference between two groups in frequencies of refraining must always remain ambiguous in this way.

Perceptually Based Indication of Labeling Tendencies (Test With Brightnesses)

Table 3 concerns the frequencies of choice of Label 1, Label 2, and of the "Equal" button during the test with brightnesses. With regard to choice of Label 1, we find Hypothesis 2*b*, that $X\bar{X} > XY$, is supported at the .001 level for Brightnesses 13–16, but Hypothesis 1*b* is not supported for those brightnesses, and neither hypothesis is supported for other brightnesses. For choice of Label 2, we find that Hypothesis 1*a*, predicting that $X\bar{X} < XY$, is supported at the .001 level for Brightnesses 9–12 and 13–16, while Hypothesis 2*a*, pre-

TABLE 3

MEDIAN FREQUENCY OF CHOICE OF LABELS 1 AND 2 AND THE REFRAIN BUTTON (R) DURING THE TEST WITH BRIGHTNESSES

Brightness	Group $X\bar{X}$ ($N = 23$)			Group XY ($N = 23$)			Group $\bar{Y}Y$ ($N = 24$)		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4, median	8.0	0.0	0.0	8.0	0.0	0.0	7.5	0.0	0.0
5–8, median	6.0	0.0	2.0	6.0	0.0	2.0	4.5	1.0	2.0
5–8, mean		0.1			0.7			1.0	
9–12, median	2.0	0.0	5.0	2.0	3.0	3.0	1.0	4.0	2.5
13–16, median	1.0	0.0	6.0	0.0	7.0	1.0	0.0	8.0	0.0
1–16, median	16.0	0.0	15.0	16.0	11.0	7.0	12.5	12.0	7.0
1–16, mean	17.1			14.4			13.2		

Brightness	Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney U Tests: Between Groups $X\bar{X}$ and XY			Mann-Whitney U Tests: Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4	N.S.	N.S.	N.S.						
5–8	N.S.	$\leq .01$	N.S.		$\leq .01$			N.S.	
9–12	N.S.	$\leq .001$	$\leq .001$		$\leq .001$	$\leq .001$		$\leq .01$	N.S.
13–16	$\leq .001$	$\leq .001$	$\leq .001$	$\leq .001$	$\leq .001$	$\leq .001$	N.S.	$\leq .02$	$\leq .06$
1–16	$\leq .02$	$\leq .001$	$\leq .001$	$\leq .09$	$\leq .001$	$\leq .001$	$\leq .09$	$\leq .01$	N.S.

dicting that $XY < \bar{Y}Y$, is supported at the .01 level for 9–12, and at the .02 level for 13–16. Hypothesis 1*a* is also supported for Brightnesses 5–8 at the .01 level, and both hypotheses are supported for the brightness dimension as a whole—Hypothesis 1*a* at the .001 level, Hypothesis 2*a* at the .01 level. Concerning "equal" judgments, Hypothesis 1*a*, that $X\bar{X} > XY$, is supported at the .001 level for both 9–12 and 13–16, and this same finding infirms Hypothesis 2*b*. Also, $XY > \bar{Y}Y$ at the .06 level for Brightnesses 13–16, a result that supports Hypothesis 2*a* and infirms Hypothesis 1*b*. The lack of significant difference between XY and $\bar{Y}Y$ for 9–12 may support both Hypotheses 1*b* and 2*a*, or simply reflect their nonconfirmation in this instance. Hypothesis 1*a* is also supported by the finding that $X\bar{X} > XY$ for the whole brightness dimension, at the .001 level.

Conceptually Based Indication of Labeling Tendencies (Test Without Brightnesses)

It is of interest to compare the individual's judgments of the past relationship between *perceived* brightnesses and class labels—the test just described above—with his judgments of the past relationship between points along the brightness dimension as he *conceptualizes* it, and those labels. The test without brightnesses provides an index of the latter, and our results indicate a high degree of match between an *S*'s judgments on the two tests—and hence between his implicit (more perceptual) and

TABLE 4

MEDIAN AMOUNT OF GRAPH AREA ASSIGNED TO LABELS 1 AND 2, AND TO 50% JUDGMENTS DURING THE TEST WITHOUT BRIGHTNESSES[a]

Brightness	Group $X\bar{X}$ ($N = 23$)			Group XY ($N = 23$)			Group $\bar{Y}Y$ ($N = 24$)		
	Label 1	Label 2	50%	Label 1	Label 2	50%	Label 1	Label 2	50%
1–4	350.0	50.0	0.0	370.0	30.0	0.0	344.5	55.5	0.0
5–8	273.0	127.0	0.0	262.0	138.0	0.0	230.0	170.0	0.0
9–12	224.0	176.0	0.0	172.0	228.0	0.0	145.0	255.0	0.0
13–16	217.0	183.0	0.0	60.0	340.0	0.0	55.0	345.0	0.0
13–16, mean			54.3			8.7			6.3
1–16	1096.0	504.0	100.0	867.0	733.0	50.0	756.5	843.5	50.0

Brightness	Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney *U* Tests					
				Between Groups $X\bar{X}$ and XY			Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	50%	Label 1	Label 2	50%	Label 1	Label 2	50%
1–4	N.S.	≤.05	N.S.		N.S.			≤.01	
5–8	≤.03	≤.03	N.S.	N.S.	N.S.		≤.01	≤.01	
9–12	≤.001	≤.001	N.S.	≤.001	≤.001		≤.02	≤.02	
13–16	≤.001	≤.001	≤.01	≤.001	≤.001	≤.01	N.S.	N.S.	N.S.
1–16	≤.001	≤.001	≤.07	≤.001	≤.001		≤.001	≤.001	

[a] 1 = 1% of a single brightness' relative frequency of occurrence.

explicit (more conceptual) models as to what went on during the training series. Table 4 presents the findings concerning the test without brightnesses: For Label 1, we find that $X\bar{X} > XY$ at the .001 level for Brightnesses 9–12 and 13–16, and also at the .001 level for the brightness dimension as a whole. Since results for Labels 1 and 2 are interdependent in the test without brightnesses, the above finding is also reflected, for Label 2, in the result that $X\bar{X} < XY$ at the .001 level for 9–12, 13–16, and the whole brightness dimension. These findings may support Hypotheses 1*a*, 2*b*, or both. Analogously, that $XY > \bar{Y}Y$ at the .02 level for Label 1, Brightnesses 9–12, is reflected also in $XY < \bar{Y}Y$ at the .02 level for Label 2, over the same brightness range—results which may support Hypotheses 1*b*, 2*a*, or both. (This effect is also present at the .001 level over the whole brightness range.) Amounts of graph area for XY and $\bar{Y}Y$ are not different, however, for Label 1 or 2 over Brightnesses 13–16, which is contrary to Hypotheses 1*b* and 2*a*. Regarding the amount of graph area assigned to "equal" judgments, that $X\bar{X} > XY$ at the .01 level supports Hypothesis 1*a* and infirms Hypothesis 2*b*. The absence of other significant differences for 9–12 and 13–16 may support all the hypotheses in question, or simply be due to their nonconfirmation in these instances.

Indifference Points

Concerning the relative positions of the median indifference points, that $X\bar{X} > XY$ at the .01 level for the test with brightnesses and for the test without brightnesses (Table 5) confirms Hypothesis 1*a* and infirms Hypothesis 2*b* in both cases; and that $XY > \bar{Y}Y$ at the .01 level for the test without brightnesses confirms Hypothesis 2*a* and infirms Hypothesis 1*b*. The nonsignificant differences between XY and $\bar{Y}Y$ for the test with brightnesses, and between $X\bar{X}$ and XY, and XY and $\bar{Y}Y$ for the training series, again may support all the relevant hypotheses or be an effect of their nonconfirmation.

Reaction Times, Training Series, and Test Without Brightnesses

If the reaction time measure is a valid index of judgment confidence, then latencies for predictions of Label 1 should drop as

TABLE 5

MEDIAN INDIFFERENCE POINT ON THE BRIGHTNESS DIMENSION[a]

Part of Experiment	Groups			Kruskal-Wallis Analysis of Variance by Ranks	Mann-Whitney *U* Tests	
	$X\bar{X}$	XY	$\bar{Y}Y$		Between Groups $X\bar{X}$ and XY	Between Groups XY and $\bar{Y}Y$
Last 32 trials of training series, median	10.0 ($N = 21$)	11.5 ($N = 17$)	9.0 ($N = 14$)	≤.09		
Test with brightnesses, median	11.0 ($N = 23$)	8.5 ($N = 21$)	8.5 ($N = 21$)	≤.001	≤.01	N.S.
Test without brightnesses, median	11.3	8.5	8.5	≤.01	≤.01	≤.01
Test without brightnesses, mean	11.5 ($N = 22$)	9.0 ($N = 23$)	8.0 ($N = 24$)			

[a] Each individual score is the median brightness assigned to the 50% judgment category by each *S*.

one moves from Brightness 8 to Brightness 1, since the association of brightness with Label 1 becomes more and more probable for all groups.[10] Further, considering predictions of Label 2 for Brightnesses 9–16, Hypothesis 1*a* predicts a sharper drop in latency (i.e., a sharper degree of slope), as one moves from 9 to 16, for Group XY than for Group $X\bar{X}$; and Hypothesis 2*a* predicts a sharper drop in latency over this brightness range for Group $\bar{Y}Y$ than for Group XY. We wished to measure, then, the degree of slope in Label 1 reaction times from Brightnesses 5–8 to Brightnesses 1–4, and in Label 2 reaction times from Brightnesses 9–12 to 13–16. As the subject score for use in analysis of these drops in latency, a measure was needed which would be free of individual differences in absolute level of reaction time, since the absolute difference between Label 1 latencies for 5–8 and 1–4, or between Label 2 latencies for 9–12 and 13–16, would be less for *S*s whose reaction times are shorter in general. This would come about because of a "floor effect": There will be less variability in Label 1 latencies for 1–4 than for 5–8, and less variability in Label 2 latencies for 13–16 than for 9–12. As a result, faster Label 1 responders for 5–8 will show less of a slope to 1–4 than the slower responders simply because the fast people's Label 1 responses to 1–4 cannot become much quicker than their responses to 5–8, while those of the slow people can. The same holds with regard to Label 2 responders for 9–12 and 13–16.

To correct for this floor effect, it hence was necessary to make the slope score a relative one—a ratio of the amount of Label 1 latency drop from 5–8 to 1–4 divided by the total possible amount of drop; and similarly, a ratio of the amount of Label 2 latency drop from 9–12 to 13–16 divided by the total possible amount of drop.

With regard to the statistical tests used for reaction time data: The first matter of interest was to determine whether the degree of downward slope in reaction time found for each group—slope for Label 1 responses from 5–8 to 1–4, or slope for Label 2 responses from 9–12 to 13–16—was significant or not. In the case of Label 1 responses, our prediction is that there will be a significant downward slope in reaction time from Brightnesses 5–8 to Brightnesses 1–4, for all groups. With regard to Label 2 responses, we expect that there will not be a significant downward slope from 9–12 to 13–16 for Group $X\bar{X}$, but that there will be a moderately significant downward slope for Group XY, and an even more significant slope of this kind for Group $\bar{Y}Y$. In the cases where we predict a difference between latencies for one group of brightnesses and another, then, our hypothesis predicts that this slope will take a particular direction: downward from 5–8 to 1–4 for Label 1 responses, and downward from 9–12 to 13–16 for Label 2 responses. Since we are thus asking whether the slope is significantly different from zero in a certain direction, a one-tail test of the latency drops is in order. Wilcoxon matched-pairs signed-ranks tests (Siegel, 1956) were used for this purpose, and their results are entered in Table 6.

Concerning the drop in latency for Label 1 responses from 5–8 to 1–4, we find the degree of this slope to be significant for all three groups during the last 32 trials of the training series, and also significant for all groups during the test with brightnesses. The significance levels met for various groups here range from .05 to .01. We also have no special reason to expect the degrees of slope to differ among the three groups; for all conditions, confidence should be greater in applying Label 1 to Brightnesses 1–4 than to Brightnesses 5–8, since association of brightnesses with Label 1 is far from certain for 5–8, and almost certain for 1–4. To determine whether there were in fact no slope differences between groups, a Kruskal-Wallis analysis of variance by ranks was

---

[10] Slope predictions for Label 1 from Brightnesses 9–16 are difficult to make since the greatest concentration of reaction times for Label 1 is expected to be from Brightnesses 1–8, approximately. Such predictions would concern Hypotheses 1*b* and 2*a*.

TABLE 6

MEDIAN SLOPE OF REACTION TIME FOR LABEL 1 RESPONSES FROM MIDDLE OF BRIGHTNESS DIMENSION TO BRIGHTEST EXTREME, AND FOR LABEL 2 RESPONSES FROM MIDDLE OF BRIGHTNESS DIMENSION TO DARKEST EXTREME[a]

Part of Experiment	Group $X\bar{X}$		Group XY		Group $\bar{Y}Y$	
	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2
Last 32 trials of training series	.07 ($N = 23$)	.03 ($N = 15$)	.08 ($N = 23$)	.09 ($N = 22$)	.07 ($N = 24$)	.19 ($N = 23$)
Test with brightnesses	.16 ($N = 23$)	.25 ($N = 4$)	.27 ($N = 22$)	.20 ($N = 20$)	.29 ($N = 23$)	.40 ($N = 24$)

Part of Experiment	Wilcoxon Matched-Pairs Signed-Ranks Tests						Kruskal-Wallis Analysis of Variance by Ranks	
	For Group XX		For Group XY		For Group $\bar{Y}Y$			
	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2
Last 32 trials of training series	$\leq .03$	N.S.	$\leq .025$	$\leq .10$	$\leq .05$	$\leq .01$	N.S.	$= .06$[b]
Test with brightnesses	$\leq .04$	N.S.	$\leq .01$	$\leq .01$	$\leq .01$	$\leq .01$	$\leq .09$	$= .05$[c]

[a] Each individual Label 1 score is the median RT of S's Label 1 responses to Brightnesses 5–8, minus median RT of his Label 1 responses to Brightnesses 1–4, over median RT of his Label 1 responses to Brightnesses 5–8. Each individual Label 2 score is the median RT of S's Label 2 responses to Brightnesses 9–12, minus median RT of his Label 2 responses to Brightnesses 13–16, over median RT of his Label 2 responses to Brightnesses 9–12.

[b] P value for the hypothesis that the three medians lie in this order.

[c] P value for the hypothesis that group $\bar{Y}Y$'s median is greater than the others.

performed comparing the three groups during the last 32 training trials, and another was done comparing these groups during the test with brightnesses. As expected, results for neither met conventional significance levels.[11]

Turning now to the latency drop for Label 2 responses from 9–12 to 13–16, we find that for the last 32 training series trials this slope is not significant for the $X\bar{X}$ group, is at the .10 level for the XY group, and is highly significant (.01 level) for Group $\bar{Y}Y$. For the test with brightnesses, the slope again is not significant for Group $X\bar{X}$, but meets the .01 level for both XY and $\bar{Y}Y$. These trends across groups are in keeping with our hypothesis, but the further task remains of determining whether indeed the slope for $\bar{Y}Y$ exceeds that for XY, and that for XY exceeds the $X\bar{X}$ group's slope, as predicted. Since we are predicting that the distributions of the slope scores will fall in a specific order for the three groups, a test by Whitney's extension of the U statistic to three groups is called for.[12] Following this procedure, our hypothesis that the slope for Group $\bar{Y}Y$ exceeds that for XY, while XY's slope exceeds $X\bar{X}$'s slope, is significant at the .06 level for the last 32 training series trials. On the test with brightnesses, the number of *S*s in Group $X\bar{X}$ providing reaction to at least one brightness in the range 9–12 and also to at least one in the range 13–16, was so small that the hypothesis of a specific order of degrees of slope for the three groups could not be tested as such. But another version of Whitney's extension of the *U* technique could be applied—one testing the related hypothesis that the slope for Group $\bar{Y}Y$ is greater than that for the other two groups. This prediction was confirmed at the .05 level. On the whole, then, the reaction time data support our expectations quite well.[13]

---

[11] That results for the test with brightnesses are at the .09 level is probably due to chance, since no comparable trend occurred for the last 32 trials of the training series. We would expect a genuine trend to be reflected in both training and test series, since the trend for Label 2 on the test with brightnesses is, as we shall see later, quite consistent with the Label 2 trend for the last 32 training series trials.

DISCUSSION

Our Hypotheses in the Light of the Results

Our hypotheses concerning acceptance limits for placing events in a class may be grouped in two ways. Some concern definition of a class of interest by inclusion or exclusion (namely, Hypotheses 1*a* and 1*b*), while others concern definition of a contrast class by inclusion or exclusion (Hypotheses 2*a* and 2*b*). One may also note that some concern the above differences when the class of interest is ambiguous (Hypotheses 1*a* and 2*a*), while others involve cases where the class of interest is partly certain (Hypotheses 1*b* and 2*b*). We find from Table 7 that Hypotheses 1*a*, 2*a*, and 2*b* tend to be confirmed more strongly than 1*b*. The appropriate conclusions would therefore seem to be these: Whether the class of interest is defined positively or negatively, and whether the contrast class is defined positively or by exclusion of the class of interest, both influence the range of acceptance limits for placing events in the class of interest, but the effect of the latter factor is somewhat greater than that of the former. Further, these effects are somewhat stronger when the class of interest is fully ambiguous than when it is partially certain, although present to a considerable extent even in the latter case.

TABLE 7

EVALUATION OF THE FOUR HYPOTHESES IN TERMS OF SOME OF THE RESULTS FOR BRIGHTNESSES 9–16[a]

Confirmations	Hypotheses			
	1*a*	1*b*	2*a*	2*b*
C/N, $P \leq .10$	7/8	2/6	6/8	4/6
C/N, $P \leq .05$	5/8	1/6	4/8	4/6

[a] We have arbitrarily assumed that each chance considered for confirmation of an hypothesis should be given equal weight with any other, and show, for each hypothesis, the number of tests that confirm it (C) divided by the total number of possibilities for its confirmation (N). Results for refraining, "equal" judgments, and indifference points are not considered here since a finding of equality between two groups on those measures may confirm two hypotheses or, alternatively, may confirm neither of them. In the first row, we define confirmations as two-group comparisons yielding *P*'s of .10 or better; in the second row, .05 or better.

---

[12] This test can be used in the present case, whereas it couldn't be applied to Tables 2–5, because the number of intergroup ties is minimal here. In dealing with those few ties that did occur, we followed Jonckheere's (1954) recommendation, which is the most conservative policy possible: namely, assigning the ties in a manner that favors the null hypothesis and is unfavorable to the hypothesis of interest to us. Specifically, ties are excluded from the sum when a large sum is favorable to our hypothesis, and ties are included in the sum when a small sum is favorable to our hypothesis.

[13] It is of interest to consider Cartwright's (1941a, 1941b) work on decision time in the light of our reaction time findings for Label 2 responses. Cartwright suggested the generalization that decision time decreases as one moves away from the boundary and toward the center of a category along some stimulus dimension. Our findings with regard to Label 2 responses suggest, however, that this is not invariably the case, but rather depends, in ambiguous situations, on whether the category in question is defined positively or negatively, and whether its contrast class is defined positively or as the complementary negative. For when the class of interest was defined negatively with a positive contrast class, we found no decrease in decision time as the center of the class was approached from its boundary; when that class was positively defined with a positive contrast class, we found some decrease; and when that class was positively defined and the contrast class was its negative complementary, we found the greatest decrease of all.

Why do these effects occur? Let us consider each of our two basic hypotheses in turn.

On Definition by Inclusion and Exclusion

A reminder is in order concerning the nature of our experimental conditions: The implications as to positively and negatively defined classes for our groups were provided by allusion, rather than by differences in logical structure of the classes. This was done so that the three groups would be equated with regard to the kinds and directness of information they were obtaining, and would differ only in whether a class *seemed* to be defined in its own right, or rather by exclusion of some other class. If the *S*s had been dealing differently with these two kinds of classes in their prior everyday experience (for whatever reason), we would thus expect such differences to transfer to the experimental situation, even though there was no advantage in the experiment proper from treating the two kinds of classes differently. Had there been such an advantage in the present experiment, we would never know if differences found between groups were due solely and specifically to the nature of this advantage or rather were due to the overwhelming influence of class definition by inclusion vs. exclusion as such. Various factors active prior to the time of the experiment could, in turn, cause a preference for positive classes—but our basic question was: Does a preference for inclusion-defined classes exist that is strong enough to operate even when it does not give *S* an advantage in the present experimental situation?

The results of our research indeed suggest that people have a prejudice of this kind in favor of classes defined by inclusion. There are several factors that may underlie this preference.

Disjunctiveness of an exclusion class. To assign an event to an exclusion class is virtually a declaration of ignorance as to how to deal with it, since the number of possible alternative actions excluded by such assignment is usually much smaller than the number of alternatives that remain. If something is not a giraffe, for example, it may be a lion, or a man, or a dandelion, or any one of a myriad of remaining possibilities. The appropriate modes of conduct toward it if it is lion, man, or anything else, are quite different from one another, and we have no basis for choosing one over another. This is the point which seems to underlie Bartlett's (1957) discussion when he notes that transfer of past learning concerning a class is possible when we are presented with new positive exemplars of that class, but not when faced with new negative exemplars of it. A negative instance of a class simply informs you that behavior appropriate to members of that class is not appropriate here, but it does not tell you what constitutes suitable behavior in the present case. The same matter is adumbrated by Hovland and Weiss (1953) when they make the following comment: "The positive instances define a coherent, well-structured combination of characteristics of the concept, while the negative instances are in effect positive instances of what the concept 'is not.' They thus lack an organized, unitary quality, except for the limiting situation where only a single dimension is involved" (p. 181).

Our study, of course, concerns just this limiting situation, since only one dimension —brightness—was used. Although a class that we defined by exclusion was thus just as organized and unitary as one we defined by inclusion, our *S*s dealt with exclusion classes as they might if their definition were disorganized, disjunctive. Such an effect may well be due to transfer from the many cases a person has encountered where exclusion classes' definitions indeed are disjunctive.

Necessity for transforming information when working with an exclusion class. Even if there is actually only one possible alternative class to which an event can belong when one knows that it is not an instance of some other class, the classifier now must usually engage in the added mental operation of assigning the event to that alternative class. ("This is not-X; therefore it is such-and-such.") There is thus one more thing he must do than would be the case if he had been able to place the event in its appropriate class at the outset. That individuals dislike this extra operation is evident from the preference for "direct test" found by Bruner, Goodnow, and Austin (1956) in their *S*s.

The extent of this preference is well illustrated by comparing particular groups from a study by Bruner, Wallach, and Galanter (in press), and work by Goodnow and Pettigrew (1956). In both cases, *S*'s task was to learn that the correct pattern of pressing a left- and a right-hand button was left-left-right repeated. In the Bruner, Wallach, and Galanter experiment *S* was informed, on any trial, which of the buttons was correct: If the left button were correct on a particular trial, a left-hand light would go on after *S* had pushed either button; while if the right button were correct on that trial, a right-hand light would appear after *S* had pressed either button. In Goodnow and Pettigrew's study, on the other hand, *S* was forced to transform information concerning which button was not correct, on trials in which *S* erred, into knowledge of what button was correct for that trial: If *S* responded correctly on any trial, a poker chip fell out of a central chute; while if he made an error on a trial, no chip appeared. When he pushed the wrong button on a trial, then, *S* had to reason that this trial was not appropriately dealt with by that button (this trial in the sequence was not an instance of the class defined by that button), but rather should be handled by the button on the other side. In contrast, when the *S* in Bruner, Wallach, and Galanter's experiment pushed the wrong button, he immediately knew which button was appropriate for that trial (i.e., he knew to which class—which button—this trial in the sequence should be assigned), since a light went on over the correct button. Positive classes thus were always utilized in the Bruner, Wallach, and Galanter study, since *S* always knew what button was appropriate on that trial. Goodnow and Pettigrew, on the other hand, provided their *S*s with complementary negative classes when they made a mistake—the *E*s told them directly only what button was *not* appropriate on that trial. That Goodnow and Pettigrew's *S*s took about twice as long to learn the left-left-right pattern as Bruner, Wallach, and Galanter's *S*s, indicates how strong an aversion people have against transforming knowledge of what something isn't into knowledge of what it is.

In our experiment, once again, such transformation is not required of *S* when dealing with an exclusion class, since its negative nature is alluded to in the label and nothing more. Placing brightnesses in the exclusion class is no less direct than placing them in the inclusion class. Nevertheless, our *S*s avoided exclusion classes as they might if there were added mental strain involved in their use. Such an effect again may be the result of transfer from the many cases a person has encountered where use of exclusion classes indeed does involve an extra step of reasoning.

Difference in kind of response. Presentation of an instance of an inclusion class requires one to make a particular suitable response of some sort, while presentation of an exemplar of an exclusion class requires one to refrain from making, or withhold, the response that would be appropriate for an instance of the excluded class. Now it may well be that the tendency to make a response is stronger, for the same amount of training, than the tendency to withhold that response; that is, given a certain amount of practice in making or withholding a response, the former tendency may be stronger than the latter. This would lead to a greater inclination to use inclusion classes rather than exclusion classes simply because of the different kinds of responses involved in each.

In our research, again, such a response difference is not obviously involved, since there is no difference in the formal arrangements of button-pushing responses for any of the three conditions: i.e., the $X\bar{X}$ and $\bar{Y}Y$ groups, as well as the XY group, all press one button to indicate placement of a brightness in Class 1, and all press an opposite button to indicate placement of a brightness in Class 2. The *S* indicates that he is using an exclusion class by as active a response as he makes to indicate use of an inclusion class. Further, one cannot argue that, whereas the verbal labels "X" and "Y" are of equal strength, the label "Yes" is stronger than "No." For our best index of the strength of "Yes" and "No," the Thorndike-Lorge (1944) frequency count, indicates that they are both very strong—and, if anything, the frequency of "No" exceeds that of "Yes." Although no obvious differences in kind of response for inclusion vs. exclusion classes were involved in our experiment, then, the *S*s reacted as they might if such response differences did occur. This once more may be due to transfer from cases where use of inclusion vs. exclusion classes did involve the sort of response differences described.

Innate preference for presence rather than absence of stimulation. We have just noted three reasons why a person might well be expected to learn, during his lifetime, to avoid exclusion classes. Apart from such learning, however, there may also be some innate basis for this preference of inclusion classes. Several theories of motivation—notably those of Freud (1955), Hull (1943), and Miller and Dollard (1941)—have conceived of drives as strong stimuli, and reduction of such stimulation intensity as a preferred state which organisms will strive to achieve. Recently, however, it has been noted that many situations exist in which organisms prefer stimulation increases rather than decreases. For example, Harlow, Harlow, and Meyer (1950) report that monkeys solve mechanical puzzles for the sheer love of doing it; Butler (1953) found that monkeys will solve problems for the "reward" of being able to look out of their cages; Kish (1955) reports that mere onset of illumination will function to strengthen a response; and evidence from the McGill group (Bexton, Heron, & Scott, 1954) indicates that mental life rapidly deteriorates in a human temporarily deprived of stimulus input through the main modalities. Thus, the presence of certain kinds of stimulation may be innately preferred to their absence. If there were a tendency of this sort at the perceptual level, it could well be expected to underlie a preference for positive rather than negative classes at the conceptual level, since a positive class involves the possible presence of its defining properties, while a negative class involves the absence of those properties and the presence of no others in particular.

Several suggestions have been made, then, as to why people prefer positive to negative classes. Any or all of them may actually apply.

On the Relevance of a Contrast Class to a Class of Interest

Our experiment has found that whether a contrast class is defined positively or by exclusion of the class of interest influences the acceptance limits for that latter class. The particular effects exercised by positive vs. negative definition of the contrast class can well be expected from the considerations outlined above, if one assumes that a contrast class will have any influence at all on the class of interest. But it is this latter question that concerns us here. Why should a contrast class influence the acceptance limits for a class of interest in the first place?

Perhaps the answer is somewhat as follows. We have noted elsewhere (Wallach, 1958) how necessary it is for the organism to select some attributes as a basis for establishing equivalence classes and to ignore others, in order to avoid being overwhelmed by the confusing richness of properties that the environment offers the senses. Attention to a contrast class would seem to be one important mechanism for making such selections, since choosing one or another contrast class tends to influence the number and kinds of properties in whose terms we define the class of interest. Further evidence on this point is contained in a recent study by Wallach and Caron (in press).

Although, once again, consideration of the contrast class does not have such an effect in the present experiment since only one dimension—brightness—was used, the demonstrated influence of the contrast class may be due to transfer from the many cases a person has encountered where contrast does serve the function discussed.

SUMMARY AND CONCLUSIONS

This study was concerned with some of the factors arising from the nature of the classification process itself that may influence our range of acceptance limits for placing events in a category. When it is uncertain whether some event should be placed in a particular class or not—whether this person is a communist, whether that student's grade should be a C, whether there was a speck of light on the telescope's screen—what factors, arising from the way classes are defined, sway our judgment? Two such factors were examined: (*a*) Placing data in a class, we proposed, does not go on in a vacuum, but rather requires that one consider some contrasting class whose nature is used as a basis for deciding how one will specify the class of interest. In particular, we expected people's readiness to assign data to the latter class to be greater when the contrast class was simply defined by exclusion as whatever remained outside the class of interest, than when defined by inclusion as being such-and-such. (*b*) Further, we noted that the class of interest itself may be defined by inclusion or exclusion—that is, in terms of its being such-and-such, or in terms of its not being some other class—and we suggested that people would be more likely to place events in the class of interest when it was defined in the former way.

Seventy university students were distributed among groups designed to test our two hypotheses under each of two conditions of ambiguity of class. It was found that both hypotheses were confirmed: that is, whether the class of interest is defined by inclusion or exclusion, and whether the contrast class is defined by inclusion or by exclusion of the class of interest, both influence the range of acceptance limits for placing events in the class of interest. However, the effect of the latter factor may be somewhat greater than that of the former. Further, these effects were found to be somewhat stronger when the class of interest was highly ambiguous than when it was partially certain, although present to a considerable extent even in the latter case.

Is it significant that these hypotheses were confirmed? Perhaps the best way to answer is to consider their origin. It occurred to us that many matters which psychologists of cognition investigate involve the organism's detecting similarities among events—involve its acting on the assumption that a present event is like something else that happened before. We thought further that such finding of similarities might best be conceptualized as a classification process. If so, judgments as to the range of equivalence—the range of events considered to be similar to one another and hence in a common class—ought to depend on the way the assumed classes are defined. Our hypotheses, in turn, concerned particular differences in such class definitions—differences which were objectively slight in the experiment but which should, on our theory, have large effects. Confirmation of these hypotheses was hence by no means a foregone conclusion, and suggested it may be valuable to explore further implications of the idea that much of cognitive activity involves classification processes.

REFERENCES

BARTLETT, F. C. *Thinking: An experimental and social study.* London: Allen and Unwin, 1957.

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BRUNER, J. S., WALLACH, M. A., & GALANTER, E. H. The identification of recurrent regularity. *Amer. J. Psychol.*, in press.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.

CARTWRIGHT, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442. (a)

CARTWRIGHT, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196. (b)

FREUD, S. Beyond the pleasure principle. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud.* Vol. 18. London: Hogarth Press, 1955. Pp. 7–64.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

HARLOW, H. F., HARLOW, M. K. & MEYER, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.*, 1950, **40**, 228–234.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

KISH, G. B. Learning when the onset of illumination is used as reinforcing stimulus. *J. comp. physiol. Psychol.*, 1955, **48**, 261–264.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289–334.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.

WALLACH, M. A. On psychological similarity. *Psychol. Rev.*, 1958, **65**, 103–116.

WALLACH, M. A., & CARON, A. J. Attribute criteriality and sex-linked conservatism as determinants of psychological similarity. *J. abnorm. soc. Psychol.*, in press.

(Accepted for publication July 16, 1958)

Vol. 73, No. 9 Whole No. 479, 1959

Psychological Monographs: General and Applied

A FACTOR-ANALYTIC STUDY ACROSS THE DOMAINS OF REASONING, CREATIVITY, AND EVALUATION[1]

NORMAN W. KETTNER,[2] J. P. GUILFORD, AND PAUL R. CHRISTENSEN[3]

University of Southern California

THIS REPORT pertains to the sixth major study in a series of investigations designed to explore abilities considered to be important in the successful performance of high-level personnel. Previous studies had been concentrated on the areas of reasoning (Green, Guilford, Christensen, & Comrey, 1953; Guilford, Christensen, Kettner, Green, & Hertzka, 1954), creative thinking (Wilson, Guilford, Christensen, & Lewis, 1954), and evaluation (Hertzka, Guilford, Christensen, & Berger, 1954). The study reported here was designed to consolidate and extend the findings of earlier studies, the three main objectives being: (*a*) verification of factors, some of which had been found only once previously; (*b*) clarification of the nature of certain factors; and (*c*) derivation of information leading to improvement of tests measuring the factors.

General Plan of the Study

The following major steps were involved in this study, in keeping with others undertaken in the project:

1. Formulation of specific hypotheses concerning the factors involved in the area of thinking investigated and concerning their properties. Only testable hypotheses were seriously entertained.
2. Selection, revision, and construction of tests that were believed to be suitable for testing the hypotheses.
3. Administration of the tests to personnel homogeneous as to sex and relatively homogeneous as to age, education, and general intellectual level.
4. Factor analysis of the intercorrelations among the tests.
5. Revision of original hypotheses, where needed, and formulations of new ones.

Selection of Factors to be Investigated

In the previous studies in the domains of reasoning, creativity, and evaluation, 31 factors had been found. Eleven of these were reference factors, which have not been regarded as belonging within these domains. Of the remaining 20 factors, 11 were chosen for further investigation in this study. The selection was determined by several considerations. The foremost reason for inclusion was the need for further information, particularly where prospects of obtaining new information seemed promising. Some factors were omitted because they were too weak in previous studies to be identified. Others were omitted because it was planned to investigate them more intensively in other studies than would have been possible in this one.

The factors selected from the previous reasoning studies were *eduction of perceptual relations, eduction of conceptual relations, eduction of conceptual patterns, eduction of correlates,* and *symbol manipulation.* Factors from the creativity study were *sensitivity to problems, associational fluency, orginality,* and *redefinition.* Factors from the evaluation study were *facility with verbal relations* and *verbal classification.*

---

[1] Under Contract N6onr-23810 with the Office of Naval Research. The ideas expressed in this paper are our own and do not necessarily represent the opinions of the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government. Of the authors of this paper, J. P. Guilford is responsible investigator and director of the project, Paul R. Christensen was assistant director, and Norman W. Kettner was in direct charge of this particular study. For more detailed information see Guilford, Kettner, and Christensen (1954, 1956).

[2] Now with Planning Research Corporation.

[3] Now with System Development Corporation.

In addition, the following reference factors were included: *general reasoning, ideational fluency, logical evaluation, verbal comprehension, numerical facility,* and *judgment.*

HYPOTHESES

Testing Hypotheses by Factor Analysis

Six of the 11 factors selected for investigation had appeared in only one previous study in this series. These factors needed to be verified; they are: *eduction of conceptual patterns, associational fluency, originality, sensitivity to problems, facility with verbal relations,* and *verbal classification.* For each of these factors, in other words, the hypothesis that the factor exists was being tested. This was done by putting in the battery tests that were loaded on the factor in a previous study together with new and presumably improved tests designed for the factor, then seeing whether the factor would emerge again.

Other hypotheses concerned the nature of each factor. For 8 of the 11 factors alternative hypotheses were formulated. For most of the alternative hypotheses at least two tests each were constructed with the expectations that they would help to make decisions between alternatives where that is possible. If the tests that were aimed at one alternative hypothesis have substantial loadings on a factor and the tests aimed at a second hypothesis do not, we may define the factor as having the properties of the first hypothesis. If tests of two or more alternative hypotheses are all substantially loaded on a factor, the definition would need to include all the properties implied. In some instances, two alternative hypotheses may actually represent two different common factors, in which case two factors would probably replace the one formerly recognized.

It will be seen in what follows that in one or two instances the same alternative hypothesis is stated pertaining to two different factors. In this case there is suspicion that the two factors, previously found in different analyses, may be one and the same. In such cases we were testing whether two such factors are, in fact, probably identical.

Some General Background of the Hypotheses

In general, alternative hypotheses were based upon consideration of results of previous studies. One alternative hypothesis for each factor was based upon the latest definition of that factor. Others might be other previously tested hypotheses that could not be rejected, or previously stated hypotheses not heretofore adequately tested. Still others represent new thinking regarding factors.

Some of the hypotheses pertain to the degree of generality of factors. In our first and only acquaintance with the factor of *sensitivity to problems,* for example, the ability seemed to be much more restricted in scope than had been originally expected. The factor had been expected to apply to the sensing of almost anything in the environment that is unusual, awry, challenging, or bothersome. As things turned out, it did not reflect anything beyond appreciating defects and deficiencies in mechanical and social devices.

The factor of *originality,* on the other hand, proved to be more general than had been expected. Tests requiring diverse evidences of *originality* remained together in defining the factor. Some were scored for uncommonness or unconventionality of responses; one was scored for cleverness of responses; and others were scored for responses showing remote associations or far-fetched ideas in connection with stimuli. Both the sensitivity-to-problems factor and the originality factor were being reexamined for degree of generality.

Other hypotheses pertain to the question of identity of pairs of factors found in different analyses. *Associational fluency,* found in the creativity analysis (Wilson et al., 1954), might be the same factor as the one called *facility with verbal relations* found in the evaluation analysis (Hertzka et al., 1954). *Redefinition* showed some possibility of identification with the

Air Force factor of *judgment* (Guilford & Lacey, 1947) in the creativity analysis but failed to show this linkage in the evaluation analysis. These points needed clarification and were expected to receive some clarification in this study.

Two factors found in the evaluation analysis were given special attention in this study—*facility with verbal relations* and *verbal classification.* There was some doubt whether these should be regarded as evaluative abilities at all. The first of these was found related to a relatively long list of verbal tests. It might even have been identified as the factor of *verbal comprehension,* except for the fact that another factor qualified for that identification. The leading test for *facility with verbal relations* was Controlled Associations, which had been included to serve as a marker for the factor of *associational fluency.* The obtained factor might have been a confounding of *associational fluency* with *eduction of conceptual relations,* the latter then regarded as a reasoning ability. In the study being reported we should have the opportunity to test these different interpretations, and, in fact, to see whether the same factor would emerge as such a second time.

There was also doubt as to the status of *verbal classification.* In the first reasoning analysis (Green et al., 1953), we entertained the hypothesis that there might be a separate ability to see identical elements in verbal entities. The common observation that classes are formed and class concepts are born would lend reasonableness to such a hypothesis. The hypothesis was proposed in connection with a group of ideas about inductive reasoning. Not enough classifying tests were included in the first reasoning analysis to determine whether there is a classifying ability distinct from an ability to see relations. So far as we know, then, both kinds of tests might be loaded with the same factor—*eduction of conceptual relations.* Thus, the factor called *verbal classification* in the evaluation analysis might have been *eduction of conceptual relations,* or it might have represented a new factor of ability to classify meaningful material. This study was expected to determine the status of the classifying-ability factor; whether or not it is separate from the relations-educing factor.

A general logical problem arose in this study owing to the fact that we introduced a number of tests that involved the act of naming. In previous studies of tests calling for the discovery of relations, patterns, and classes, we had generally asked for multiple-choice responses indicating that these had been discovered from the fact that the examinee did something revealing. This is an indirect way for the examinee to show that he has cognized the relation, pattern, or class. The additional step or steps might well bring additional common factors into the test variances. By asking the examinee to name or describe the relation, pattern, or class, it was thought that we should obtain a more direct indication of successful cognitions.

It soon occurred to us that naming, itself, is an additional act and might also generate additional common factors. Another aspect of this problem would be whether, if the act of naming does entail something new in the way of factors, there would be one new naming factor, in common to all of the naming tests, or there would be several naming factors, each parallel to one of the discovery factors—that for discovering perceptual relations, that for conceptual relations, and that for conceptual patterns—and perhaps, also, parallel to other factors. Considering all the naming tests included in the battery, there were possibilities for all the way from none to as many as five naming, or verbalizing, factors to appear.

Statement of the Hypotheses

1. *Eduction of perceptual relations* had been found in tests composed of figural material and not in tests composed of meaningful words. Two kinds of figural tests were involved: those involving seeing simple relationships as to form, position, shading, and other properties and others involving the classification of figures according to their properties. We therefore tested the hypothesis that there is a single ability involving both relations and classes

of figural material or that there are two hypothetical factors: (*a*) seeing figural relations and (*b*) seeing figural classes. A hypothesis tested incidentally was whether tests involving the grouping of letters and spelling of words would also be related to one or both factors.

2. *Eduction of conceptual relations* had had relatively low loadings in previous analyses, but there was little doubt of its existence as a factor. As in the case of the preceding factor, however, there was the question of whether one factor would suffice or whether it would take two: (*a*) seeing conceptual relations and (*b*) seeing conceptual similarities or seeing classes.

3. *Eduction of conceptual patterns* seemed to call for something more than simple relationships. The term "patterns" had been used because it is sufficiently general to include principles, systems, rules, and perhaps trends. It had been designated as "conceptual" at the time this study was planned because the patterns were based upon combinations of elements in which figural properties were of little or no importance. Three alternative hypotheses were entertained in this study: (*a*) seeing patterns, (*b*) seeing rules or principles, and (*c*) seeing trends.

4. *Eduction of correlates* was defined as the ability to find something that fits into a given relationship. Both conceptual and nonconceptual tests had been previously loaded on the factor. One important question was whether, if given the opportunity, conceptual and nonconceptual tests would split into two groups, indicating two distinct but parallel abilities, as we had seen in a number of places. The leading test for the factor previously had included items involving relations between meanings and items involving relations between word structure or spelling. In the present study the two kinds of items were put in separate tests. The two alternative hypotheses were: (*a*) finding something to fit a perceptual relationship and (*b*) finding something to fit a conceptual relationship.

5. *Symbol manipulation* was defined as the ability to manipulate symbols according to given rules. In another analysis a factor called "symbol substitution" was defined as the ability to substitute symbols according to rules. We had the question as to whether there is one factor or two. With added tests included in this study we had the opportunity to answer this question and also to determine the main properties of either one or two factors. The hypotheses were: (*a*) symbol manipulation and (*b*) symbol substitution.

6. *Verbalizing ability* was suggested as a new factor (or group of factors), in accordance with the earlier discussion. We had eight tests in the battery that involve the naming or describing of discovered relations, of similarities or identical elements, of rules or principles, and of trends. We thought that there might be as many as five factors and we have accordingly five hypotheses: (*a*) naming perceptual relations, (*b*) naming conceptual relations, (*c*) naming classes, (*d*) stating rules or principles and (*e*) describing trends. Either *c* or [illegible] might divide two ways, depending upon the test content, whether figural or verbal.

7. *Sensitivity to problems* had been found limited to two tests, involving mechanical devices and social institutions, respectively, in the creativity analysis. Tests were added to determine whether the factor generalizes at all to nonmechanical and nonsocial areas. The only hypothesis tested here is with regard to the generality of the factor.

8. *Associational fluency* had been defined as the ability to produce words to meet specific requirements of meaning. The test Controlled Association, which called for 12 synonymous responses to each item, strongly determined this factor in the creativity analysis. Another form of the same test, which called for only three responses to each item, failed to segregate the same factor in the evaluation analysis. Both forms of the test were included in the present study. We also had two tests that called for only one word per item. The two hypotheses were: (*a*) producing one or two words to fit a given meaning and (*b*) producing several words to fit a given meaning.

It might be found that a single factor would have both properties or two separate factors might be found.

9. *Originality* was being tested for degree of generality, as stated earlier. Originality tests had been constructed along three lines, involving the production of uncommon, remote, or clever responses, respectively. All three kinds of tests had been loaded on the factor, but in this study tests were added in ways that should have given the opportunity for two or even three originality factors to emerge. The three hypotheses were: (*a*) production of remote responses, (*b*) production of uncommon responses, and (*c*) production of clever responses.

10. *Redefinition* had been a rather weak factor in two prior analyses. It was defined as the ability to shift the function of an object or part of an object and use it in a new way. In this study we sought to determine whether the factor could be verified and whether it could be distinguished from the *judgment* factor. As things turned out, conditions were not favorable for answering the second question.

11. *Facility with verbal relations,* as indicated earlier, was a factor of uncertain status. In this study we were merely interested in determining whether tests previously defining the factor would hang together as before or whether they would define two or more other factors, such as *associational fluency* and *eduction of conceptual relations.*

12. *Verbal classification,* found in one of our previous analyses, offered two logical possibilities. The factor was defined as the ability to appraise poorly defined verbal relationships. Here relation to a category or class is actually implied. The "poorly" was added because tests of the same character but having more clear-cut, logical relationships did not have loadings on the factor. An apparently similar factor found by Adkins and Lyerly (1951) was defined as an ability to perceive abstract similarities. The two hypotheses are: (*a*) appraising poorly defined verbal relationships and (*b*) seeing conceptual similarities.

Summary of the Hypotheses and Reference Factors

For convenience, the various hypotheses just stated are presented in outline form to serve as ready references in connection with later discussion of tests and factors:

1. Eduction of perceptual relations
    (*a*) Seeing perceptual relations
    (*b*) Seeing perceptual similarities
2. Eduction of conceptual relations
    (*a*) Seeing conceptual relations
    (*b*) Seeing conceptual similarities
3. Eduction of conceptual patterns
    (*a*) Seeing conceptual patterns
    (*b*) Seeing rules or principles
    (*c*) Seeing trends
4. Eduction of correlates
    (*a*) Finding something to fit a given perceptual relation
    (*b*) Finding something to fit a given conceptual relation
5. Symbol manipulation
    (*a*) Symbol manipulation
    (*b*) Symbol substitution
6. Verbalizing ability
    (*a*) Naming perceptual relations
    (*b*) Naming conceptual relations
    (*c*) Naming classes
    (*d*) Stating rules or principles
    (*e*) Describing trends
7. Sensitivity to problems
8. Associational fluency
    (*a*) Producing one or two words to fit a given meaning
    (*b*) Producing several words to fit a given meaning
9. Originality
    (*a*) Producing remote responses
    (*b*) Producing uncommon responses
    (*c*) Producing clever responses
10. Redefinition
11. Facility with verbal relations
12. Verbal classification
    (*a*) Appraising poorly defined verbal relations
    (*b*) Seeing conceptual similarities

Since many tests designed to test the hypotheses were known to contain or were suspected of containing common variances with factors outside this list, reference tests were added to the battery to isolate these sources of variance. The reference factors, numbered in sequence with those already listed, are:

13. Ideational fluency
14. General reasoning
15. Logical evaluation

16. Verbal comprehension
17. Numerical facility
18. Judgment

THE TEST BATTERY

The hypotheses were the main consideration in determining the kinds of tests that were selected for inclusion in this study. In those cases in which tests had to be adapted or new tests had to be constructed to fit the hypotheses, several other considerations were kept in mind. Each test was divided into at least two separately timed, comparable parts for purposes of convenient and meaningful estimation of reliability. Variance due to personal experiences of examinees was minimized as much as possible. An attempt was made to keep the factorial complexity of tests as low as possible. Before the assembly of the battery, new or adapted tests were pretested on samples of college students. On the basis of these results, 57 experimental tests were selected for inclusion in the analysis. It had been planned to include several of the Air Force tests, which are given to air cadets routinely for classification purposes, to serve as markers for some of the reference factors. But by the time our battery could be administered, the Air Force had discontinued separate scoring of its classification tests, hence the test scores were not available to us.

The tests are described in the following section. For convenience the tests are listed in alphabetical order. For every test the nature of the task is described and for most a sample item is given. Other information includes the number of separately timed parts, the number of items per part, and the working time. The code numbers of the tests indicate the following information. The first letter, R, C, E, or P, indicates reasoning, creativity, evaluation, or planning, respectively. The second letter (or pair of letters) stands for a specific hypothesis. Next comes the number of the test in that category, 01, 02, etc. The last letter indicates the form of the test, A meaning the first form, B the second, and so on. The uncoded tests were taken over from batteries given before the coding system was instituted.

Description of Tests[4]

1. *Apparatus Test—CS01B.* A list of 20 implements is presented. The *E*'s (examinee's) task is to suggest two improvements for each of them. Sample item:

Suggest two improvements for the telephone.

1. *A device to record messages when no one is at home.*
2. *Luminous dials to operate in the dark.*

Parts, 2. Items per part, 10. Working time, 14 min. hfc,[5] 7.

2. *Associations III—CN05A.* The items consist of pairs of words. The *E*'s task is to think of a single word that is similar in meaning to both the given words. Since the given words are not synonyms, the word that *E* supplies must have a different meaning in relation to each of the given words.

Sample item:

recline __*lie*__ deceive

The correct answer is "lie" since it has a meaning similar to both given words.

Parts, 2. Items per part, 15. Working time, 12 min. hfc, 8*a*, 9*a*, 11.

3. *Associations IV—CN06A.* This test is made up of pairs of words like those in Associations III. In this test, however, *E*'s task is to think of a word that is associated with (not synonymous with) the given words and that has a different meaning in its relation to each of the given words.

Sample item:

jewelry __*ring*__ bell

A correct association word is "ring." It has a different meaning in its relation to each of the given words. It is a piece of jewelry and also the sound of a bell.

---

[4] The authors gratefully acknowledge permission for test use and/or adaptation granted by the following authors or publishers: D. C. Adkins and D. E. Baier (Tests 31 and 56); R. I. Blakey (Tests 5 and 18); L. L. Thurstone (Tests 9, 10, 23, and 51); National Council for Social Studies (Test 12); and the Ronald Press Company (Test 25). Tests 13 and 44 were developed by the U. S. Air Force and Test 17 was developed by the U. S. Navy. Acknowledgment is also made for permission granted by the *Saturday Evening Post* and *Collier's* to use their cartoons in Test 4. We also wish to thank Constance D. Lovell for making available to us a number of classes in beginning psychology for preliminary experimental testing.

[5] hfc stands for hypothesized factor content. See preceding summary for key to the hypotheses.

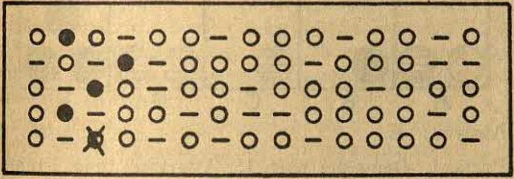
Parts, 2. Items per part, 15. Working time, 14 min. hfc, 9*c*.

4. Cartoons—CN07A. Each item is a cartoon taken from old issues of *Saturday Evening Post* or *Collier's*. The *E*'s task is to supply clever "punch" lines for each of the cartoons.

Parts, 2. Items per part, 8. Working time, 14 min. hfc, 9*c*.

5. Circle Reasoning (no code number). Each item consists of five rows of circles and dashes. One of the circles in each of the first four rows has been blackened according to a rule. The *E*'s task is to find the rule and then mark the circle that should be blackened in the last row.

Sample item:

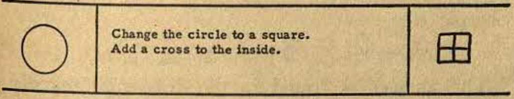


The second circle from the left in each row should be black; therefore, the second circle in the last row is marked to indicate that it is the one which should be black.

Parts, 1. Items per part, 15. Working time, 8 min. hfc, 3*a*.

6. Completion of Figural Changes—REC05A. In each item a figure is given. Written instructions as to how the figure changes are also given. The *E*'s task is to draw a figure that shows the results of all the changes.

Sample item:



If the circle is changed to a square and a cross is put into the square, then the figure that is drawn at the right is the correct answer.

Parts, 2. Items per part, 10. Working time, 10 min. hfc, 4*a*.

7. Consequences Test—CF06A (low-quality score). The *E* is to write what would happen if certain changes were suddenly to take place. He is to list as many different consequences or results of these changes as he can.

Sample item:

What would happen if all national and local laws were suddenly abolished?

1. *No more parking tickets*
2. *There would be more killing.*
3. *No more taxes*

Two scores are derived from this test. The low-quality score is the number of direct responses, i.e., those responses that describe direct or immediate consequences of the given changes.

Parts, 4. Items per part, 1. Working time, 8 min. hfc, 13.

8. Consequences Test—CF06A (remoteness score). The remoteness score is the second score that is derived from the Consequences Test. It is the number of remote responses, i.e., those responses that describe remote or far-reaching consequences of the given changes.

Parts, 4. Items per part, 1. Working time, 8 min. hfc, 9*a*.

9. Controlled Associations—CAF01B. In each item *E* is given a common word. His task is to write as many synonyms as possible for each word, 12 spaces being provided for each word.

Sample item:

GOOD: *right* *virtuous* *deserving*
excellent *noble* *pleasing*
superior *worthy* *creditable*
commendable *pious* *admirable*

Parts, 2. Items per part, 4. Working time, 14 min. hfc, 8*b*.

10. Controlled Associations II—CAF02B. The format of this test is the same as Controlled Associations, CAF01B, except that *E* may give only three synonyms for each given word.

Sample item:

ODD: *unusual*
unique
queer

Parts, 2. Items per part, 10. Working time, 6 min. hfc, 8*a*, 8*b*, 11.

11. Correlate Completion II—REC01A. Each item consists of two pairs of words and a single word. The *E*'s task is to supply a word that bears the same relation to the single word as the relation between the words in the first two pairs. The relation between the words in each pair is based on the structure of the words and not on their meanings.

Sample items:

1. am ma not ton tool *loot*
2. enrage rage correlate late about *bout*

Parts, 2. Items per part, 20. Working time, 12 min. hfc, 4*a*.

12. Critical Evaluation—EL05A. The statements of this test have been taken from the section of newspapers entitled "Letters to the Editor." It is *E*'s task to designate those statements (A) based primarily on emotion or prejudice, and those (B) showing reasoning and thought.

Sample items:

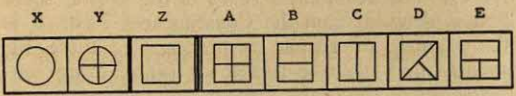
1. All people who drink hard liquor should have their driver's licenses taken away.
2. The police should revoke the licenses of those people who drive while drunk.

Answers: 1. A, 2. B.

Parts, 2. Items per part, 15. Working time, 5 min. hfc, 12*a*.

13. Figure Analogies—RPR04A. The *E*'s task is to select a figure that bears the same relation to the third figure as the second bears to the first.

Sample item:

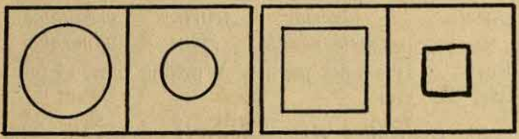


Answer: A.

Parts, 2. Items per part, 15. Working time, 10 min. hfc, 1*a*, 4*a*.

14. Figure Analogies Completion (no code number). This test is the same as Figure Analogies RPR04A except that *E* must draw the figure that completes the second pair in accordance with the relation given by the first pair.

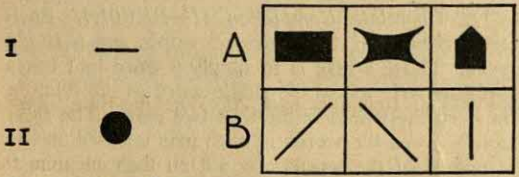
Sample item:



Parts, 1. Items per part, 15. Working time, 10 min. hfc, 4*a*.

15. Figure Classification—EF01A. Classes are defined by groups of three figures. Single figures have to be assigned to these classes in terms of common characteristics. There are three sets of five classes and five single figures have to be classified for each set.

Sample item:

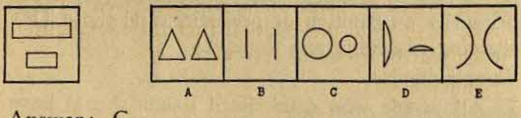


Answer: I - B, II - A.

Parts, 1. Items per part, 15. Working time, 6 min. hfc, 1*b*.

16. Figure Matching—EF04A. In each item a geometrical figure is given along with a set of five other figures. The *E*'s task is to select the figure from the set that has most in common with the key figure.

Sample item:

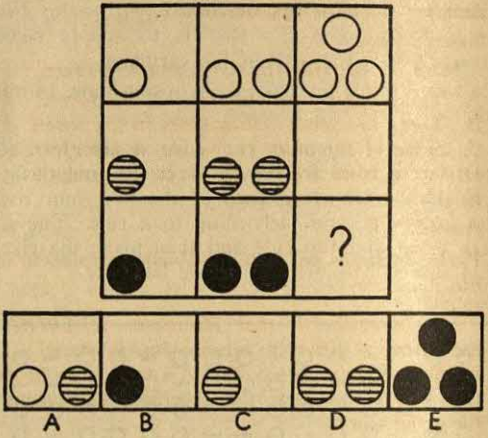


Answer: C.

Parts, 2. Items per part, 15. Working time, 14 min. hfc, 1*b*.

17. Figure Matrix (no code number). In each item nine squares are presented with figures that change in both directions. Some squares are left empty. The *E* indicates which one of five alternate figures goes in an indicated square.

Sample item:



Answer: E.

Parts, 1. Items per part, 15. Working time, 7 min. hfc, 1*a*.

18. Form Reasoning (no code number). In each item three familiar forms are given. In addition, a table showing how these forms can be combined is given. On the basis of this table, *E* is to find a form that is equal to the three given forms.

Table:

☆ ○ = □	○ ☆ = □	○ ✙ = ☆	✙ ○ = ☆
□ ✙ = △	✙ □ = △	☾ ✙ = ◇	✙ ☾ = ◇
☆ ✙ = ○	✙ ☆ = ○	◇ ○ = ✙	○ ◇ = ✙
□ ○ = ☾	○ □ = ☾	○ ☾ = 8	☾ ○ = 8

Sample item:

☆ ○ ✙ = ☆ □ ☾ △ ✙

The answer is found in the following manner. According to the table above, the first two figures—that is, the star and circle—are equal to a square. Now, if a square is put in place of the star and circle, then a square and a cross remain. According to the table a square and a cross are equal to a triangle. Therefore, the correct answer is a triangle.

Parts, 1. Items per part, 20. Working time, 2 min. hfc, 5*b*.

19. Gestalt Transformation—CR03A. In each item a problem is presented. The *E*'s task is to select from five given objects the one that would be most useful in solving the problem.

Sample item: TO START A FIRE

A. fountain pen B. onion C. pocket watch D. light bulb E. bowling ball

Best answer: C.

Parts, 2. Items per part, 10. Working time, 10 min. hfc, 10.

20. Inference Test—RD03B. Each item presents a statement of fact or opinion, such as might be found in a newspaper, and five alternate conclusions. The *E* indicates which of these can be drawn with most justification.

Sample item:

Most of the trees in the forest are green.

A. There are no yellow trees in the forest.
B. There are some yellow trees in the forest.
C. Some of the trees in the forest are green.
D. Green trees are the tallest in the forest.
E. Pine trees are green.

Answer: C.

Parts, 2. Items per part, 8. Working time, 10 min. hfc, 15.

21. Inventive Verbal Relations—REC07A. In each item a different relationship is stated. Two words are then given. The *E*'s task is to produce two words that bear the specified relationship to the given words.

Sample item:

(a) is the opposite of (b)

(a) black: (b) *white* :: (a) strong: (b) *weak*

Parts, 2. Items per part, 15. Working time, 8 min. hfc, 4*b*.

22. Letter Grouping—RPS01A. In each item four groups of four letters each are given. Three of the groups are alike in some way. The *E*'s task is to pick out the group of letters that is different from the other three.

Sample item:

AABC ACAD ACFH AACG

Parts, 2. Items per part, 20. Working time, 8 min. hfc, 1*b*.

23. Letter Series—RCP03A. Each item consists of a series of letters. The *E*'s task is to find the rule for the series and write the next two letters in the series.

Sample item:

E F E F E F E *F* *E*

Parts, 2. Items per part, 15. Working time, 8 min. hfc, 3*b*.

24. Letter Triangle—RCP01B. Each item consists of a series of letters arranged in triangular form with various letters left out. The *E*'s task is to determine which letter goes into the blank space indicated by the question mark.

Sample item:

a A. h
b c B. j
d e f C. g
? D. f
— — — — E. a

Answer: C.

Parts, 2. Items per part, 8. Working time, 12 min. hfc, 3*a*.

25. Logical Classification—EL07A. This test consists of statements containing "ought." The *E* has to classify these statements in terms of categories, which are defined and for which examples are given.

Sample item:

You ought not to cross the street against the red light.

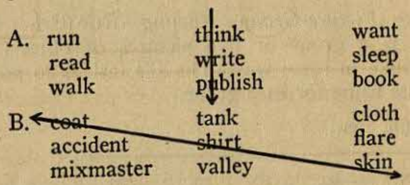
A. custom B. completeness C. safety D. utility E. welfare

Answer: C.

Parts, 2. Items per part, 25. Working time, 14 min. hfc, 10, 11, 15.

26. Matrix Order—PR01A. Each item consists of a 3 × 3 matrix of words. The *E*'s task is to draw a straight line through three words that represent a sequence in space or time. He is also to indicate the direction of the sequence by arrow heads.

Sample items:



Parts, 2. Items per part, 15. Working time, 18 min. hfc, 3*c*.

27. Number Series—RCP05A. In each item a series of numbers is given. The *E*'s task is to find the rule for the series and write a description of it in the given blank.

Sample item:

15 18 21 24 27 30 *+3*

Parts, 2. Items per part, 20. Working time, 12 min. hfc, 3*b*, 6*d*.

28. Object Synthesis—CY01B. In each item the names of two objects are given. The *E*'s task is to think of something that could be made by combining the two objects.

Sample item:

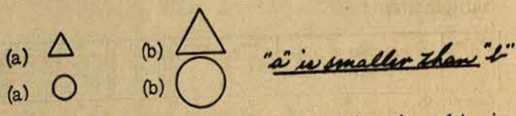
GIVEN: volley ball steel spring

Possible answer: punching bag.

Parts, 2. Items per part, 12. Working time, 10 min. hfc, 10.

29. Perceptual Relations Naming—RPR03A. In each item two pairs of figures are given. The *E*'s task is to state the relation that the first figure of each pair bears to the second.

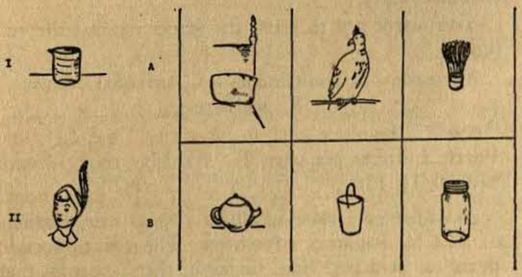
Sample item:



Parts, 2. Items per part, 14. Working time, 14 min. hfc, 1*a*, 6*a*.

30. Picture Classification—EF03A. Classes are defined by groups of three pictures each. The *E*'s task is to assign single pictures to these classes in terms of common characteristics.

Sample item:

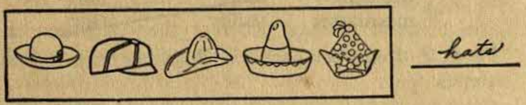


Answers: I - B, II - A.

Parts, 2. Items per part, 15. Working time, 13 min. hfc, *2b, 12b.*

31. Picture-Group Naming—RCS03A. Each item is a group of five pictures of objects that are alike in some way. The *E*'s task is to provide a class name for each group.

Sample item:



Parts, 1. Items per part, 24. Working time, 6 min. hfc, *2b, 6c, 12b.*

32. Plot Titles—CF02C (cleverness score). In the Plot Titles test, four story plots are presented. The *E* is instructed to write as many clever titles as he can for each plot. The cleverness score is the number of responses that are rated clever.

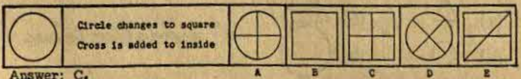
Parts, 4. Items per part, 1. Working time, 12 min. hfc, *9c*

33. Plot Titles—CF02C (low-quality score). The low-quality score is the number of responses that are rated "not clever."

Parts, 4. Items per part, 1. Working time, 12 min. hfc, 13.

34. Prescribed Relations—RSM03B. In each item one figure is given. Written instructions as to how the figure is to be changed are also given. The *E*'s task is to pick out from five given alternatives the figure that shows the results of all the changes.

Sample item:



Answer: C.

Parts, 2. Items per part, 10. Working time, 10 min. hfc, *4a.*

35. Quick Response—CN01A. This test resembles the conventional word-association test. A list of words is read to the examinees at the rate of one every five seconds. The *E* is instructed to write down the first word that comes to him.

The test is scored by assigning a weight to each response where weights are directly proportional to infrequency of occurrence in the population, and summing the weights.

Parts, 1. Items per part, 50. Working time, 4.2 min. hfc, *9b.*

36. Remote Verbal Similarities—RCS02B. In this test *E*'s task is to select a word from five alternatives that has most in common with a given word, where all connections are rather remote.

Sample item:

FATHER
A. candidate
B. second baseman
C. agitator
D. superintendent
E. salesman

Answer: D.

Parts, 2. Items per part, 15. Working time, 10 min. hfc, *9a, 12a.*

37. Seeing Deficiencies—PJ01A. In each item a short description of a plan or activity is given. The *E*'s task is to point out how the plan is faulty.

Sample item:

A growing city discovers pressing needs to improve both its streets and its sewer systems. After due consideration, the council decides to work on the street-improvement program first. What is wrong with this plan?

The streets will have to be torn up again for the sewer system.

Parts, 2. Items per part, 10. Working time, 20 min. hfc, 7.

38. Seeing Problems—CS06A. In each item the name of a common object is given. The *E*'s task is to write down five problems that might arise in connection with the object.

Sample item:

CANDLE: A. *How to light it.*
B. *Keeping it from falling.*
C. *Keeping it from flickering.*
D. *How long will it burn.*
E. *What to do with the drippings.*

Parts, 2. Items per part, 6. Working time, 8 min. hfc, 7, 13.

39. Seeing Trends—RI06A. The items in this test consist of a group of words that tend to change progressively in some meaningful way. The *E*'s task is to discover the trend in the words.

Sample item:

Mouse rat lion pig cow horse elephant
Animals become larger.

Parts, 2. Items per part, 10. Working time, 10 min. hfc, 3*c*, 6*e*.

40. Seeing Trends II—RCP04A. This test is similar to Seeing Trends RI06A. Each item consists of a group of words and *E*'s task is to find the trend in the words. In this test, however, the trend is based on the structure of the words rather than on the meaning of the words.

Sample item:

Anger bacteria camel dead excite
The first letters go from "a" to "e"

Parts, 2. Items per part, 12. Working time, 16 min. hfc, 3*c*, 6*e*.

41. Ship Destination (no code number). Using a diagram and observing an increasing number of rules, *E* has to determine the distance between ships and ports. In addition to rules of distance, wind direction must be considered at the start of the test. Later the current direction, the strength of wind and current, and the heading of the ship must all be taken into account.

Parts, 1. Items per part, 48. Working time, 15 min. hfc, 14.

42. Sign Changes—CX04B. This test involves simple numerical operations in which *E* must substitute one arithmetic operation sign for another before performing the operation.

Sample item:

In the following problems wherever you see:

$-$ replace it by $\times$
$+$ replace it by $-$
$3 - 6 = \underline{18}$
$6 + 2 = \underline{\ 4\ }$

Parts, 3. Items per part, 8, 16, 16. Working time, 1.67 min. hfc, 5*b*.

43. Sign Changes II—RSM02A. The items consist of simple arithmetic equations. In these equations the value on the left of the "equals" sign is not equal to the value on the right. The *E*'s task is to indicate which interchange of signs will make the two values equal.

Sample item:

$1 + 2 = 4 \times 1$

A. $+$ and $-$
B. $\times$ and $+$
C. $+$ and $\div$
D. $\times$ and $\div$
E. $\times$ and $-$

Answer: E.

Parts, 2. Items per part, 10. Working time, 14 min. hfc, 5*a*.

44. Similarities—RCS01A. Each item consists of a pair of common objects. The *E*'s task is to write six ways in which the objects are alike.

Sample item:

Apple and Orange are alike

A. *sweet*
B. *round*
C. *have seeds*
D. *fruit*
E. *grow on trees*
F. *have skins*

Parts, 2. Items per part, 6. Working time, 10 min. hfc, 2*b*, 12*b*, 13.

45. Social Institutions—CS05A. In each item the name of a social institution is given. The *E* is asked to list two improvements for each social institution that is given.

Sample item:

Marriage:

Courses on marriage in schools.
Social approval of women proposing.

Parts, 2. Items per part, 6. Working time, 10 min. hfc, 7.

46. Syllogisms I—EL01A. The items in this test represent 15 combinations of the four figures and four moods of the categorical syllogism. The *E* is required to state whether the conclusion is true or false in each case.

Sample item:

All soldiers are men.
Some citizens are soldiers.
Therefore, some citizens are men.

Answer: True.

Parts, 2. Items per part, 15. Working time, 10 min. hfc, 15.

47. Symbol Manipulation II—RSM04A. In each item a main statement, presented symbolically, is given along with several conclusions. Each conclusion consists of two parts. The *E*'s task is to indicate the truth or falsity of each conclusion on the basis of given definitions of the symbols.

Sample item:

"L" means "larger than"
"S" means "smaller than"
"E" means "equal to"
"NE" means "not equal to"

If xEy and yNEz, then:

1. xNEz and yEx
2. xSz and xLz

A. Both parts are true.
B. One part is true and the other part is indeterminate.
C. One part is true and the other part is false.
D. One part is false and the other part is indeterminate.
E. Both parts are false.

Answers: 1 - A, 2 - C.

Parts, 2. Items per part, 15. Working time, 10 min. hfc, 5*a*, 15.

48. Unusual Uses—CN04B. The items are names of common objects. Each object has a common use, which is given. The *E* is asked to list other uses for which the object or parts of the object could serve.

Sample item:

A newspaper (used for reading)

1. *to start a fire.*
2. *to wrap garbage in.*
3. *to swat flies.*
4. *stuffing to pack boxes.*
5. *to line drawers or shelves.*
6. *to make up a kidnap note.*

Parts, 2. Items per part, 3. Working time, 10 min. hfc, 9*b*, 13.

49. Verbal Analogies I—RCR01C. This is a verbal-analogies test in which there was an attempt to make the relationship between the first pair of words relatively difficult to grasp and the analogy relatively easy to fulfill.

Sample item:

electricity: energy :: green: ______?

A. color B. red C. grass D. trees E. foliage

Answer: A.

Parts, 2. Items per part, 15. Working time, 12 min. hfc, 2*a*.

50. Verbal Analogies Completion—REC08A. This is a completion form of the verbal-analogies test.

Sample item:

cloth: dye :: house: *paint*

Parts, 2. Items per part, 25. Working time, 8 min. hfc, 4*b*.

51. Verbal Classification—EL16A. Vertical groups of eight words are flanked on either side by vertical groups of four words that define two classes. After deciding upon the nature of the defined classes, *E* goes down the center column of words, checking on the left if a word belongs in the class to the left, checking on the right if it belongs in the one on the right, and making no mark if the word does not belong in either class.

Sample item:

COW		desk	√	TABLE
HORSE	√	sheep		CHAIR
GOAT		rocker	√	BOOKCASE
DOG		tree		LAMP
	√	cat		
		nose		
		dresser	√	
	√	donkey		

Parts, 2. Items per part, 40 (5 groups of 8). Working time, 8 min. hfc, 2*b*, 12*b*.

52. Verbal Comprehension (no code number). This test is Part I of the Guilford-Zimmerman Aptitude Survey (short form). It is a standard multiple-choice vocabulary test.

Parts, 1. Items per part, 40. Working time, 12 min. hfc, 16.

53. Verbal Relations Naming—RCR02A. Each item consists of two pairs of words. The *E*'s task is to write down the relation that the first word in each pair bears to the second.

Sample item:

(a) dwarf (b) giant

(a) bottle (b) barrel (a) *is smaller than* (b)

Parts, 2. Items per part, 25. Working time, 12 min. hfc, 2*a*, 6*b*.

54. Vocabulary Completion—REC06A. The *E*'s task is to write down a word that fits a given definition and that begins with a given letter.

Sample item:

A contest of speed..........(r) *race*

Parts, 2. Items per part, 40. Working time, 6 min. hfc, 8*a*, 16.

55. Word Classification—EL17A. The *E*'s task is to decide which word in a group of four does not belong.

Sample item:

A. horse B. cow C. man D. flower

Answer: D.

Parts, 2. Items per part, 20. Working time, 14 min. hfc, 2*b*, 12*b*.

56. Word-Group Naming—RCS02B. Each item consists of groups of five words. The *E*'s task is to write down a definitive name for each group of words.

Sample item:

knife pan bowl rolling pin strainer

cooking utensils

Parts, 2. Items per part, 15. Working time, 10 min. hfc, 2*b*, 6*c*, 12*b*.

57. Word Groups—RPS02A. Each item consists of a group of four words. The *E*'s task is to find out what the words in each group have in common. The things that the words have in common are structural rather than meaningful.

Sample item:

read retire rearming restless

All words begin with "re"

Parts, 2. Items per part, 15. Working time, 20 min. hfc, 1*b*, 3*b*, 6*c*.

TESTING CONDITIONS

It had been originally planned that this battery of tests would be administered to a single sample. Arrangements had been made in December of 1953 to administer the battery to aircrew trainees at Lackland Air Force Base, San Antonio, Texas. The total testing time amounted to 12 hours.

During the interval between the time the arrangements were made and the test material was shipped, conditions had changed at Lackland Air Force Base. Twelve hours of testing time could not be worked into the schedule of the trainees and a maximum of six hours was offered. The solution was to divide the battery into three smaller ones, Batteries A, B, and C, each taking six hours. It was later learned that the Air Force could not work Battery B into its schedule so arrangements were made to have it administered to naval air cadets at Pensacola, Florida.

An overlapping of the batteries enabled us still to test most of the hypotheses, although with some loss of information. There was something gained by the fact that some factors could come out in more than one analysis, hence there was some opportunity for replication of results. Two tests, Nos. 20 and 46, had to be omitted entirely, which meant loss of two tests for the factor of *logical evaluation.*

Test Administration

In setting up the batteries, special attention was given to the order of administration of the tests, seeing that the tests designed for a certain hypothesis were not given consecutively. Administration manuals provided complete information to test administrators. From both testing locations it was reported that no special problems arose during the test administration and that the examinees seemed well motivated.

The Samples

Batteries A and C were administered at Lackland Air Force Base to 205 and 215 aircrew trainees, respectively, in groups of approximately 100 each, during the month of August, 1954. Battery B was administered to 225 examinees in groups of 30 to 60 every two weeks during the months of October through December, 1954.[6] At both locations the aviation cadets were typically in their early 20's, above the average of the population as to general intelligence, and had education of the level of high-school graduation or beyond.

Some of the examinees had to be eliminated from the study because of incomplete information regarding them. As a result, the samples numbered 201 for Battery A, 219 for Battery B, and 210 for Battery C.

TREATMENT OF THE DATA

Scoring

The responses to 20 of the tests were recorded on IBM answer sheets by the examinees. The answer sheets were machine scored and the scoring checked. All scores involved a correction for guessing. The remaining tests being of completion form were scored by hand and the scoring checked.

Test Statistics

Reliabilities were estimated using a sample of 200. Alternate forms and odd-even estimates were made for 49 of the tests. For two tests the Kuder-Richardson Formula 20 was used. Other estimates were obtained from earlier studies. The various estimates of reliability and their sources are given in Table A,[7] along with means and standard deviations of the various test variables.

The Factor-Analysis Procedures

After checking upon score distributions, which all approached the normal form, Pearson product-moment correlation coefficients were computed among the tests for the three batteries. These data are presented in Tables B, C, and D.

Twelve factors were extracted from each of the correlation matrices, using Thurstone's complete centroid method. The loadings of the last centroid factors were all smaller than .20 in absolute value and all residuals were less than .05 in absolute value. The unrotated factor matrices are presented in Tables E, F, and G.

The axes were rotated by Zimmerman's (1946) graphic, orthogonal method. The three criteria used to guide the rotations were simple structure, positive manifold, and psychological meaningfulness. The rotated factor matrices are presented in Tables 1, 2, and 3.

---

[6] Grateful acknowledgment is made to the Personnel Research Laboratory, HRRC, Lackland AFB, Texas, and to the Aviation Psychology Laboratory, U. S. Naval Air Station, Pensacola, Florida, for making this testing possible. In particular, the authors wish to thank Lloyd G. Humphreys and J. W. Bowles at Lackland and Wilse B. Webb at Pensacola. In addition, the authors wish to thank the military and civilian personnel who acted as test administrators and proctors.

[7] Tables hereinafter designated Tables A-G have been deposited with the American Documentation Institute. Order Document No. 5958 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

ROTATED FACTOR MATRIX—BATTERY A

	A	B	C	E	F	G	H	J	K	M	U		
	V	GR	PR	CR	VC	EP	EC	SM	SS	CC	(d)	R_1	h^2
5.	−.01	.08	.14	.13	.08	.44	.23	.01	−.06	.09	−.01	−.28	.39
6.	.29	.10	.40	.12	.05	.15	.31	.18	.17	.17	.28	.09	.57
11.	.04	.17	.11	.17	.24	.25	.42	.17	.17	.31	.19	.05	.56
12.	.39	.18	−.09	.06	.10	.03	−.02	−.04	.29	.04	−.01	−.27	.37
13.	.18	.27	.54	.16	.15	.29	.25	.07	.08	−.02	.05	.14	.63
14.	.12	.18	.33	.17	−.01	.07	.53	−.02	−.02	−.03	.26	.19	.58
16.	.13	−.12	.00	−.10	.13	.26	.04	−.12	.14	.02	.09	.28	.25
17.	.12	.19	.39	−.03	.13	.27	−.01	.03	−.10	.14	.09	.10	.34
18.	.01	.25	.20	.15	−.08	.16	.08	.02	.53	.07	−.04	.03	.45
21.	.41	.09	.14	.23	.12	.11	.18	.14	.23	.51	.13	−.04	.66
23.	.08	.41	.14	.15	.18	.36	.40	.15	.14	.27	.00	.04	.66
24.	.11	.08	.17	.20	−.02	.38	.33	.26	.02	.09	.07	.03	.42
26.	.42	.02	.05	.03	.10	.08	.21	.19	−.04	.20	.07	.10	.33
27.	.02	.36	.11	.04	.22	.32	.16	.43	.32	.20	−.03	.15	.67
29.	.03	.29	.45	.13	.11	.17	.31	.09	.16	.30	.01	−.07	.57
30.	.28	−.05	.28	.37	.15	.17	.25	.20	.09	.00	−.06	−.20	.50
34.	.07	.26	.27	.29	.21	.14	.12	.20	.07	.01	.41	.08	.53
39.	.17	.10	−.04	.19	.00	.21	.08	−.06	.05	.29	.59	.00	.57
40.	.09	.11	.10	.29	.24	.11	.31	.12	.26	.28	.14	.17	.49
41.	.04	.61	.06	.18	.00	.07	.20	.07	.03	.06	.01	.03	.46
42.	.08	.22	−.01	.12	−.01	−.01	.21	.19	.65	.05	.04	−.26	.64
43.	−.06	.37	.11	.14	.06	.07	.28	.39	.09	.07	.21	−.06	.47
47.	.08	.23	.16	.13	.24	.08	.12	.43	.08	.10	.26	.07	.45
49.	.49	.23	.16	.48	.04	−.02	.21	.13	.02	.04	.19	.21	.69
50.	.38	.12	.28	.38	.29	.16	.13	.02	−.05	.35	.15	.00	.66
51.	.14	.21	.13	.30	.32	.09	.14	.06	.07	.29	.19	−.05	.43
53.	.28	.19	.24	.29	.04	.16	.13	.18	.11	.54	.01	−.09	.64
55.	.42	.09	.04	.20	.45	−.01	.28	.03	.05	.05	.09	.26	.59

TABLE 2

Rotated Factor Matrix—Battery B

	A V	C PR	D PC	E CR	F VC	I SR	L NA	M CC	P AF	R_2	R_3	R_4	h^2
2.	.31	.12	.09	.04	.19	.16	.20	.18	.46	.20	−.01	−.04	.51
3.	.12	.11	.20	.13	.16	−.09	.17	.02	.40	.21	.27	−.15	.45
10.	.28	−.04	.05	.16	.27	.12	.14	.03	.48	−.13	.00	.05	.47
11.	.14	.20	.25	.22	.10	.50	.06	.44	.27	−.01	.06	.03	.71
12.	.37	−.06	−.12	.14	−.01	.12	.28	−.07	.11	−.10	.19	.09	.34
13.	.10	.58	.18	.28	.05	.21	.18	.19	−.04	−.11	.08	−.16	.62
15.	.18	.30	.37	.04	.10	.00	.06	.14	.14	−.05	−.13	.04	.33
16.	.08	.28	.17	.00	.12	.14	−.10	.07	.00	−.28	.06	−.13	.26
17.	.17	.55	.13	−.02	−.08	.17	.13	.07	−.04	.04	−.03	.11	.42
21.	.53	.08	.19	.05	.11	.04	.29	.47	.20	−.08	.24	.00	.75
22.	−.07	.25	.29	.07	.17	.56	.27	.14	.24	−.12	.03	−.05	.67
29.	.29	.19	.06	−.07	.15	.39	.22	.34	−.01	.00	.02	−.08	.47
30.	.15	.02	.40	.24	.10	.24	.22	−.09	.08	.06	−.15	−.05	.40
31.	.12	−.08	.33	.05	.07	.08	.38	.21	.15	−.01	.15	.15	.40
36.	.32	.15	.25	.09	.12	−.05	.01	−.05	.14	−.02	.29	.08	.33
39.	.29	.03	−.07	−.03	.19	.14	.34	.00	.02	−.08	−.20	−.07	.31
40.	.20	.15	.04	.15	.17	.42	.08	.19	.23	−.03	.09	.12	.41
49.	.39	.14	.28	.43	.13	.28	.00	.07	−.09	−.05	.08	−.01	.55
50.	.43	.22	.09	.37	.15	.08	.22	.41	.17	−.05	.04	−.05	.66
51.	.20	.23	.18	.34	.40	.04	.19	.20	.18	−.02	−.02	.03	.51
52.	.79	.05	.01	.15	.08	.12	−.03	.11	.10	−.19	.05	.09	.74
53.	.28	.12	−.04	.10	.28	.18	.32	.33	.16	−.02	.11	.20	.50
54.	.40	−.09	.20	.11	.16	.20	.23	.44	.41	.04	−.12	−.06	.72
55.	.39	.00	.29	.20	.40	.09	−.10	.13	.00	.05	.21	−.01	.52
56.	.31	−.07	.17	.29	.38	.04	.36	.13	.00	−.12	.05	−.02	.52
57.	.22	.11	.28	.22	.29	.38	−.09	.19	.27	.12	.09	−.17	.58

TABLE 3

ROTATED FACTOR MATRIX—BATTERY C

	A	I	L	N	O	P	Q	R	S	T			
	V	SR	NA	SP	Pe	AF	O	Re	IF	SX	R_5	R_6	h^2
1.	−.11	−.03	.05	.45	.51	.20	.17	−.03	.24	.22	.01	.03	.65
2.	.42	.00	.21	−.05	−.09	.36	.12	.35	−.08	.20	.12	−.20	.60
3.	.29	.15	.24	−.06	−.08	.47	.02	.24	.00	−.04	.07	−.24	.52
4.	.18	−.05	.10	.04	.27	.18	.47	.25	.16	−.05	.00	.20	.50
7.	.07	.08	−.10	.09	.24	.03	.06	.06	.58	.24	−.09	.08	.50
8.	.16	.29	.08	.23	.12	.22	.32	−.08	.34	.41	.03	−.09	.63
9.	.27	.13	.04	.04	.28	.56	.30	.01	.19	.23	.15	.01	.69
10.	.29	.09	.02	.28	.27	.61	.19	−.05	.18	.27	.03	.13	.78
19.	.05	−.03	.11	−.06	−.08	.26	.15	.48	−.11	.11	.03	.10	.38
22.	.00	.49	.27	.28	−.07	.26	.06	.19	.09	−.02	.20	−.08	.56
25.	.39	.17	.01	.10	.10	.05	.21	.44	−.18	.11	.08	−.18	.52
28.	.11	−.02	.40	−.05	.29	.16	.26	.06	.18	.40	.09	−.08	.56
31.	.11	.17	.37	.29	.07	.28	.04	.19	−.07	.12	−.08	.02	.41
32.	.21	.25	.14	.06	.15	.16	.43	.04	−.04	.01	−.18	−.02	.40
33.	−.12	−.04	−.02	.12	.20	.04	.06	−.05	.60	.01	−.21	−.15	.51
35.	.02	−.11	−.17	.14	.01	−.01	.18	−.15	.24	−.11	.16	−.22	.26
36.	.22	.07	.29	.01	.14	.02	−.08	.34	.05	.00	.22	−.05	.33
37.	.15	.11	.20	.39	−.03	.07	.23	.40	.00	.21	.02	.17	.52
38.	.13	.13	−.04	.50	.40	.17	.27	.00	.14	.17	.05	−.27	.67
44.	.19	.10	−.01	.24	.48	.28	.19	.23	.26	.10	.10	−.06	.59
45.	.18	.06	.13	.38	.50	.16	.24	.00	.27	.10	−.20	.09	.66
48.	.03	.07	.13	.05	.48	.25	.35	.13	.14	.31	.10	−.13	.60
52.	.64	.12	.24	.05	.02	.13	.16	.13	−.14	−.04	.07	.07	.57
54.	.44	.12	.24	.18	−.05	.33	.23	.24	.20	.03	.17	.16	.62
56.	.37	.24	.46	.07	.28	.22	.04	.11	−.07	−.02	.09	−.04	.57
57.	−.04	.49	−.02	.00	.24	.17	.07	.29	−.03	.05	.09	.07	.43

INTERPRETATION OF THE FACTORS

The factors will be discussed roughly in order of the hypotheses listed earlier. The interpretations rest principally upon tests with factor loadings of .30 or greater. The names and numbers of the tests are listed preceding the discussion for each factor. Following the names of the tests are three columns labeled A, B, and C representing the three batteries. Factor loadings .30 or greater are listed opposite the name of the test in the column representing the battery in which the factor appeared. If a test had a loading of .30 or greater on a particular factor in one battery and one less than .30 in another battery on the same factor, both loadings are listed. In parentheses are loadings .30 or greater on other factors. A dash in the column indicates that the test was not in the particular battery. A plus sign indicates that the test was in the battery but that the factor under discussion did not appear in that battery.

The first two factors to be discussed are reference factors.

FACTOR A. VERBAL COMPREHENSION (V)

	A	B	C
49. Verbal Analogies I	.49 (.48 CR)	.39 (.43 CR)	—
55. Word Classification	.42 (.45 VC)	.39 (.40 VC)	—
26. Matrix Order	.42	—	—
21. Inventive Verbal Relations	.41 (.51 CC)	.53 (.47 CC)	—
12. Critical Evaluation	.39	.37	—
50. Verbal Analogies Completion	.38 (.38 CR, .35 CC)	.43 (.37 CR, .41 CC)	—
52. Verbal Comprehension	—	.79	.64
54. Vocabulary Completion	—	.40 (.44 CC, .41 AF)	.44 (.33 AF)
36. Remote Verbal Similarities	—	.32	.22 (.34 Re)
56. Word-Group Naming	—	.31 (.38 VC, .36 NA)	.37 (.46 NA)
2. Associations III	—	.31 (.46 AF)	.42 (.36 AF, .35 Re)
25. Logical Classification	—	—	.39 (.44 Re)

It was expected that some of the tests designed for this study would contain appreciable amounts of *verbal-comprehension* variance in spite of the fact that an effort was made to keep to a low vocabulary level. The relatively large number of tests in the list above indicates that this is the case.

FACTOR B. GENERAL REASONING (GR)

	A	B	C
41. Ship Destination	.61	—	—
23. Letter Series	.41 (.36 EP, .40 EC)	—	—
43. Sign Changes II	.37 (.39 SM)	—	—
27. Number Series	.36 (.32 EP, .43 SM, .32 SS)	—	—

Ship Destination was put into Battery A as a reference test for *general reasoning* which had appeared in all four of the previous studies. According to most recent information (Kettner, Guilford, & Christensen, 1956), *general reasoning* is best defined as the ability to comprehend or to structure problems in preparation for solving them.

FACTOR C. EDUCTION OF PERCEPTUAL RELATIONS (PR)

	A	B	C
13. Figure Analogies	.54	.58	—
29. Perceptual Relations Naming	.45 (.31 EC, .30 CC)	.19 (.39 SR, .34 CC)	—
6. Completion of Figural Changes	.40 (.31 EC)	—	—
17. Figure Matrix	.39	.55	—
14. Figure Analogies Completion	.33 (.53 EC)	—	—
15. Figure Classification	—	.30 (.37 PC)	—

This factor is the first in the list of factors being studied in this investigation. There were three tests listed under Sub-hypothesis 1*a* (seeing perceptual relations) —Figure Analogies, Figure Matrix, and Perceptual Relations Naming. All have significant loadings on this factor in both batteries (except Perceptual Relations Naming which has a significant loading only in Battery A). For this reason this factor is called *eduction of perceptual relations* and is defined as the ability to see perceptual relations.

Three other tests have significant loadings on the factor. One of these is Figure Analogies Completion. This test probably should have been included under Sub-hypothesis 1*a,* since the examinee must discover the relation between two figures before he draws a figure that fits the relation. As such it is reasonable to expect that a portion of the variance of this test would be accounted for by a seeing-relations factor. Another test loaded on this factor is Completion of Figural Changes. It was expected that the major portion of the variance of this test would be accounted for by an *eduction-of-correlates* factor since the relationship is given and the examinee has only to draw a figure to fit the relationship. The final test on this factor is Figure Classification which is a "seeing similarities" test. It is possible that relationships between figures have to be taken into account in classifying them.

In both reasoning studies the list of tests for *eduction of perceptual relations* had contained tests in which seeing similarities seemed to be the main feature along with tests in which seeing relations seemed to be the main feature. One hypothesis tested in this study was that each of these two types of tests would form a factor of its own if given a chance. This seems to be the case, since Factor C involves seeing perceptual relations and Factor D involves seeing perceptual similarities.

FACTOR D. PERCEPTUAL CLASSIFICATION (PC)

	A	B	C
30. Picture Classification	+	.40	—
15. Figure Classification	—	.37 (.30 PR)	—
31. Picture-Group Naming	—	.33 (.38 NA)	+

This factor is called *perceptual classification* and is defined as the ability to see similarities in perceptual material. It had been expected that two of the tests—Figure Classification and Figure Matching—under Sub-hypothesis 1*b*(seeing perceptual similarities), would define the factor if it emerged. Figure Matching had low intercorrelations with all other tests and as a result has no significant loading on any factor in this study. Figure Classification does have a significant loading on the factor.

Two other tests were listed under Sub-hypothesis 1*b*. They are tests of the type that have been called "structural" in this project. Structural tests are defined as those that involve simple symbols, letters, or numbers. The hypothesis that "seeing similarities" in structural material is the same as "seeing similarities" in figural material was tested in this study. This hypothesis was not substantiated since neither of the structural tests has a significant loading on the factor.

The other tests with significant loadings, Picture Classification and Picture-Group Naming, were expected to have significant loadings on Factor F, *verbal classification.* The items in both these tests consist of groups of pictures of objects. It was felt that seeing similarities in items of this type would be more closely related to seeing similarities in groups of words than to seeing similarities in groups of figures. This does not seem to be the case. It may be that figural properties help to determine classes of the objects in these tests.

FACTOR E. EDUCTION OF CONCEPTUAL RELATIONS (CR)

	A	B	C
49. Verbal Analogies I	.48 (.49 V)	.43 (.39 V)	—
50. Verbal Analogies Completion	.38 (.38 V, .35 CC)	.37 (.43 V, .41 CC)	—
30. Picture Classification	.37	.24 (.40 PC)	—
51. Verbal Classification	.30 (.32 VC)	.34 (.40 VC)	—

This factor, analogous to Factor C, is in the conceptual area and is defined as the ability to see relationships in conceptual material. The first test in the list above was included under Sub-hypothesis 2*a* (seeing conceptual relations). The second test probably should have been included under that sub-hypothesis since seeing a relation should occur before a word is written to fit the relation. The other test under Sub-hypothesis 2*a*, Verbal Relations Naming, was also expected to have a significant loading on this factor. An effort was made to keep the seeing of the relations as easy as possible, however, so that the major portion of the variance would be contributed by the naming of the relations. A nonsignificant loading here indicates that this attempt was successful.

In addition to the two analogy tests there is a verbal-classification test with significant loadings in both batteries. This situation is similar to that in *eduction of perceptual relations* where there is a figural-classification test. Picture Classification also has a significant loading on the factor in one of the batteries. The fact that this test has significant loadings on both a perceptual-discovery and a conceptual-discovery factor indicates that both properties may be important in picture tests. Since Picture Classification has a significant loading in only one battery, however, this significant loading may be due to chance.

As in the case of Factor C, in spite of the fact that classification tests have significant loadings on this factor, the definition is restricted to seeing relationships since seeing similarities seems to be the unique feature of the next factor to be discussed.

FACTOR F. VERBAL CLASSIFICATION (VC)

	A	B	C
55. Word Classification	.45 (.42 V)	.40 (.39 V)	—
51. Verbal Classification	.32 (.30 CR)	.40 (.34 CR)	—
56. Word-Group Naming	—	.38 (.31 V, .36 NA)	+

This is the *verbal-classification* factor that corresponds to the *perceptual-classification* factor. All three tests in the list above were listed under Sub-hypothesis 2*b* (seeing conceptual similarities). They all involve the classification of verbal material. Two other tests under Sub-hypothesis 2*b*, Picture Classification and Picture-Group Naming, have significant loadings on the *perceptual-classification* factor (D) rather than on this one. The final test under Hypothesis 2*b*, Similarities, was not included in either Battery A or Battery B.

A factor called *verbal classification* had been found in the evaluation analysis (Hertzka et al., 1954) where it was defined as the ability to classify verbal material. The two leading tests in the evaluation analysis were Verbal Classification and Critical Evaluation. The two leading tests in the present study are Verbal Classification and Word Classification. In spite of the fact that these tests were included in both studies, the factors in the two studies have only one test in common, therefore it is not known whether the two factors are the same or whether they represent two different factors. If the two factors are distinct, the factor found in this study has a stronger claim to the name *verbal classification,* since the tests with significant loadings here seem to fit the definition better than those with significant loadings in the evaluation study.

Factor G. Eduction of Patterns (EP)

	A	B	C
5. Circle Reasoning	.44	—	—
24. Letter Triangle	.38 (.33 EC)	—	—
23. Letter Series	.36 (.41 GR, .40 EC)	—	—
27. Number Series	.32 (.36 GR, .43 SM, .32 SS)	—	—

A factor similar to this was found in the first reasoning analysis (Green et al., 1953). There it was called *eduction of conceptual patterns* and was defined as the ability to see rules, principles, or systems. All the tests loaded on the factor in this study were listed under Sub-hypotheses 3*a* and 3*b*. They all involve seeing rules or principles that are more complex than simple relations. The answers to the items in the Circle Reasoning test involve several conditions and the items in the other three tests consist of double and triple series. By contrast, the items in the trends tests under Sub-hypothesis 3*c* all involve one relation that is repeated between succeeding pairs of elements. In view of these facts, the definition of this factor was changed to the ability to see a *set* of relationships.

All the tests in the list above can be considered to be structural so that this factor might be called "eduction of structural relations." There were no figural or verbal tests in the study similar to these structural tests. Thus it is not known whether the definition of this factor should be restricted to structural tests or not. For this reason the factor is called simply *eduction of patterns* pending new empirical evidence.

Factor H. Eduction of Correlates (EC)

	A	B	C
14. Figure Analogies Completion	.53 (.33 PR)	—	—
11. Correlate Completion II	.42 (.31 CC)	+	—
23. Letter Series	.40 (.41 GR, .36 EP)	—	—
24. Letter Triangle	.33 (.38 EP)	—	—
6. Completion of Figural Changes	.31 (.40 PR)	—	—
29. Perceptual Relations Naming	.31 (.45 PR, .30 CC)	+	—
40. Seeing Trends II	.31	+	—

The two leading tests here, Figure Analogies Completion and Correlate Completion II, were included under Sub-hypothesis 4*a* (finding something to fit a given perceptual relation). Three other tests were listed under this sub-hypothesis. Two of these, Completion of Figural Changes and Prescribed Relations, are very similar. Both involve finding figures to fit a given relationship. One is a completion test, however, and the other is a multiple-choice test. The completion test, Completion of Figural Changes, has a significant loading whereas the multiple-choice test, Prescribed Relations, does not. This is not too surprising since completion tests have been prominent in relation to this factor and since this factor stresses producing something to fit a given relation. The last test under Sub-hypothesis 4*a* was Figure Analogies. The seeing-relations aspect of this test seems to be most important since it has a significant loading only on *eduction of perceptual relations*.

Two of the other tests in the list, Letter Series and Letter Triangle, are structural tests listed in connection with Hypothesis 3, which has to do with the eduction of patterns. The discovery aspect probably gives them a place in the list for Factor G. In each test the examinee has to show that he knows the system by giving other letters to extend the system. Letter Series is a completion test and, although Letter Triangle is a multiple-choice test, the extension problem is still an important aspect of the test. But the presence of these two tests in the list suggests that *correlate completion* is not only a matter of supplying something to complete a relationship; it seems also to be a matter of supplying something to complete a pattern or set of relationships.

The two remaining tests in the list are Perceptual Relations Naming and Seeing Trends II. In neither test is there any need to complete either a relationship or a system. It will be noted that these two tests were also in Battery B along with Correlate Completion II without determining a correlate-completion factor. This suggests that the two tests may not belong in the list for Factor H. Reasons for this will be suggested a little later.

One of the important reasons for including the eduction-of-correlates factor in this study was to see whether it might be replaced by two factors, one perceptual and one conceptual. We seem to have an answer to this question, for neither of the tests affiliated with Sub-hypothesis 4*b* (finding something to fit a conceptual relation) had significant loadings on this factor, while three of the four tests under Sub-hypothesis 4*a* (finding something to fit a perceptual relation) had significant loadings. The two conceptual tests largely determine Factor M in this analysis.

Factor H might be called "eduction of perceptual correlates" except for the fact that both figural and structural tests are loaded on it. For this reason it is best to call it simply *eduction of correlates.* Later research may show that it should be replaced by two factors, one for figural material and one for structural.

The presence of the two unexpected tests—Seeing Trends II and Perceptual Relations Naming—can be accounted for logically by reference to the list of tests under Factor I, to be discussed next. They are both significantly related to the factor of *eduction of structural relations,* along with Correlate Completion II. The latter factor did not come out in analysis of Battery A, as might well have been the case. What probably happened is that it became confounded with Factor H in Battery A.

FACTOR I. EDUCTION OF STRUCTURAL RELATIONS (SR)

	A	B	C
22. Letter Grouping	—	.56	.49
11. Correlate Completion II	+	.50 (.44 CC)	—
40. Seeing Trends II	+	.42	—
29. Perceptual Relations Naming	+	.39 (.34 CC)	—
57. Word Groups	—	.38	.49

All the tests in this list except Perceptual Relations Naming are structural tests that were scattered among several hypotheses. Since we have already seen that Battery B yielded a factor of *eduction of perceptual relations* and a factor of *eduction of conceptual relations,* since all but one of the tests in the list for Factor I are composed of letters, and since most tests involve seeing relations, we are forced to recognize a third relations-seeing factor and to give it an appropriate name—*eduction of structural relations.* The presence of a figural test in the list may reflect the fact that there is some correlation between this factor and the parallel figural one.

All is not otherwise satisfactory with the picture of Factor I. Two tests—Letter Grouping and Word Groups—call for the seeing of similarities rather than for seeing relations in the ordinary sense. It is interesting that these two tests determine a doublet factor in analysis of Battery C, from which the remaining tests in the list for Factor I are missing. There is a possibility that the factor in Battery C is another one, an ability to see classes based upon structural material. We have already seen that a perceptual-classification factor and a verbal-classification factor are distinguished from eduction-of-relations factors in Factors D and F. It may be that there is also a structural-classification factor distinct from a corresponding *eduction of structural relations.* If this is the case, Factor I in Battery B is a confounding of two factors. Without further evidence, however, we may for the present accept the two factors listed under Factor I as being identical. A study now under way should determine whether there is also a structural-classification factor.

FACTOR J. SYMBOL MANIPULATION (SM)

	A	B	C
47. Symbol Manipulation II	.43	—	—
27. Number Series	.43 (.36 GR, .32 EP, .32 SS)	—	—
43. Sign Changes II	.39 (.37 GR)	—	—

Both tests that were included under Sub-hypothesis 5*a*, symbol manipulation, have significant loadings on this factor. Thus, the factor is defined as the ability to manipulate symbols. Number Series was not predicted to have a significant loading but it is quite likely that manipulation of symbols is involved in working this test. Numbers, of course, are symbols and it appears that some trial-and-error manipulation is necessary before a solution can be reached. Trial-and-error manipulation seems also to be a feature of the other two tests with significant loadings, and the trial-and-error aspect probably could be added to the definition of the factor. None of the tests that were included under Sub-hypothesis 1*b* has a significant loading on this factor.

There is a possibility that this factor is the same as the *handling-complicated-procedures* factor found in the general-reasoning study (Kettner et al., 1956) or it may be a combination of *handling complicated procedures* and *symbol manipulation*. Another possibility is that this factor is a structural one parallel to *logical evaluation.*

FACTOR K. SYMBOL SUBSTITUTION (SS)

	A	B	C
42. Sign Changes	.65	—	—
18. Form Reasoning	.53	—	—
27. Number Series	.32 (.36 GR, .32 EP, .43 SM)	—	—

Sign Changes and Form Reasoning were the two tests included under Sub-hypothesis 5*b*, *symbol substitution,* so this factor can be defined as the ability to substitute symbols quickly. This factor differs from the preceding one in that no trial-and-error manipulation seems to be necessary. All that is required is the ability to substitute symbols according to simple, given rules.

As in the previous factor, Number Series has a significant loading on this factor but here the finding is not so reasonable. One possible explanation is that this factor is a combination of *symbol substitution* and *numerical facility.* Both Sign Changes and Number Series involve numbers and Form Reasoning generally has had a near-significant loading on numerical facility in previous studies. The high loading of Form Reasoning argues against the possibility that this is entirely a number factor.

FACTOR L. NAMING ABSTRACTIONS (NA)

	A	B	C
31. Picture-Group Naming	—	.38 (.33 PC)	.37
56. Word-Group Naming	—	.36 (.31 V, .38 VC)	.46 (.37 V)
39. Seeing Trends	+	.34	—
53. Verbal Relations Naming	+	.32 (.33 CC)	—
28. Object Synthesis	—	—	.40 (.40 SX)

This appears to be one of the hypothesized verbalizing factors. The first four tests in the list above were included under the general Hypothesis 6 (verbalizing ability). They call for the naming of classes, relations, or trends. The last test in the list above, Object Synthesis, calls for naming objects that fit certain requirements. On the basis of these considerations the factor has been called *naming abstractions.*

This factor is similar to one found by Adkins and Lyerly (1951), which they called *concept formation.* The first two tests in the list above led their list for the *concept-formation* factor. The factor here seems to be more restricted than the one found by Adkins and Lyerly. All tests on *naming abstractions* are verbal tests except Picture-Group Naming while the tests on *concept formation* included perceptual and structural tests as well as verbal tests. In addition, all tests that have significant loadings here are completion-type tests that require verbal answers. In the Adkins' study, both completion and multiple-choice tests had significant loadings on *concept formation.* Perhaps a portion of their *concept formation* has been segregated in this study.

A naming factor was also found by Carroll (1941). The leading tests on that factor involved speed of naming colors and speed of naming forms. It does not seem that *naming abstractions* is sufficiently similar to Carroll's naming factor to be identified with it.

FACTOR M. EDUCTION OF CONCEPTUAL CORRELATES (CC)

	A	B	C
53. Verbal Relations Naming	.54	.33 (.32 NA)	—
21. Inventive Verbal Relations	.51 (.41 V)	.47 (.53 V)	—
50. Verbal Analogies Completion	.35 (.38 V, .38 CR)	.41 (.43 V, .37 CR)	—
11. Correlate Completion II	.31 (.42 EC)	.44 (.50 SR)	—
29. Perceptual Relations Naming	.30 (.45 PR, .31 EC)	.34 (.39 SR)	—
54. Vocabulary Completion	—	.44 (.40 V, .41 AF)	+

The tests with significant loadings on this factor are scattered among various sub-hypotheses. Since both tests listed under Sub-hypothesis 4*b* (finding something to fit a given conceptual relation) have significant loadings on this factor, it might well be called *eduction of conceptual correlates.* The other tests, however, do not seem to fit this interpretation so well. They are completion-type tests, but the main task in two of them is that of naming an observed relationship. One of these tests is a figural test and one other test is a structural test.

One rationalization would be to extend the conception of the *eduction of conceptual correlates* to the act of producing words that stand for abstract ideas, the ideas being given or readily discovered. This conception brings us close to the definition of another factor—*expressional fluency*—which had been defined as the ability to put ideas into words, at the time this study was conducted. More recent information, however, fairly well indicates that *expressional fluency* is an ability to produce connected discourse; to form sentences with appropriate sentence structure. The tests in the list for Factor M call for single words as responses, or simple phrases, at the most, which fairly well eliminates the hypothesis that this is the factor of *expressional fluency.* It seems, then, that we have found a conceptual factor for the production of correlates, parallel to the previously known perceptual factor of the same kind. The presence of the naming tests may require a somewhat broader conception of the factor or it may mean some confounding with an unidentified factor.

FACTOR N. SENSITIVITY TO PROBLEMS (SP)

	A	B	C
38. Seeing Problems	—	—	.50 (.40 Pe)
1. Apparatus Test	—	—	.45 (.51 Pe)
37. Seeing Deficiencies	—	—	.39 (.40 Re)
45. Social Institutions	—	—	.38 (.50 Pe)

This factor is called *sensitivity to problems,* which has been defined as the ability to see defects, needs, and deficiencies. All tests that were hypothesized to measure *sensitivity to problems* have significant loadings on the factor here, and no other tests have significant loadings. These tests all involve the seeing of problems of some kind. Seeing Problems involves problems connected with common objects; Apparatus Test involves problems with mechanical devices; Social Institutions involve problems with social institutions; and Seeing Deficiencies involves problems with plans or activities.

In the creativity study (Wilson et al., 1954), only two of these tests, Apparatus Test and Social Institutions, were present and had significant loadings on *sensitivity to problems.* Thus the factor seemed to be rather restricted. It was included in this study in order to test its degree of generality by adding somewhat different tests. From the variety of tests in the list above it appears that the factor is more general than had been indicated in the previous study.

FACTOR O. PENETRATION (Pe)

	A	B	C
1. Apparatus Test	—	—	.51 (.45 SP)
45. Social Institutions	—	—	.50 (.38 SP)
44. Similarities	—	—	.48
48. Unusual Uses	—	—	.48 (.35 O, .31 SX)
38. Seeing Problems	—	—	.40 (.50 SP)

Since this factor has three tests in common with the preceding one, there is some difficulty in distinguishing between the two. A clue is given, however, by the two tests, Unusual Uses and Similarities, which have significant loadings on this factor but not on the previous one. Neither is strictly a seeing-problems test. The Unusual Uses test calls for listing unusual uses for a common object, and the Similarities test calls for naming similarities between common objects.

In the creativity study, Social Institutions and Apparatus Test had additional significant loadings on a doublet which was not interpreted although it was suggested that it might be a trace of a "penetration factor," which had been hypothesized. In that study, penetration was hypothesized as an ability to see beyond the immediate and the obvious.

Three tests—Apparatus Test, Seeing Problems, and Social Institutions—in the list above have significant loadings on Factor N *(seeing problems).* If it can be assumed for these three tests that the variance for seeing immediate or obvious problems is accounted for by Factor N, then Factor O might well represent the variance due to seeing problems that are not so immediate or obvious. It is to be expected that those individuals who can see beyond the immediate and obvious in the Similarities test, should be able to produce more similarities, and should be able to see far-fetched uses of objects in the Unusual Uses test.

The definition of this factor comes fairly close to Sub-hypothesis 9*a* (producing remote responses). It is quite possible that this factor corresponds to that sub-hypothesis (see discussion under Factor Q). If this is so, one question of interest here is why the tests under Sub-hypothesis 9*a* do not have significant loadings on this factor. One of these tests is Consequences (remote score). Perhaps one reason this test does not have a significant loading on Factor O is that it deals with things that are remote in time whereas the tests loaded on Factor O do not. The other three tests under Sub-hypothesis 9*a* were objectively scored tests whose items involved remote meanings of words. These tests are of a different type than those loaded on Factor O.

Factor P. Associational Fluency (AF)

		A	B	C
10.	Controlled Associations II	—	.48	.61
2.	Associations III	—	.46 (.31 V)	.36 (.42 V, .35 Re)
54.	Vocabulary Completion	—	.41 (.40 V, .44 CC)	.33 (.44 V)
3.	Associations IV	—	.40	.47
9.	Controlled Association	—	—	.56 (.30 O)

All tests with significant loadings on this factor were listed under the two sub-hypotheses under *associational fluency.* They all require the production of words, but they vary in several different ways. The stimuli are of three different kinds—one given word, two given words, or a definition. The words to be produced have five different restrictions—words synonymous with one given word, words synonymous with two given words, words fitting given definitions, words associated with one given word, and words associated with two given words. The number of correct responses that are possible for each item varies from one to many. The number of responses that are asked for in each item varies from one to twelve. Some of the tests are speed tests and the others are relatively more in the nature of power tests. None of these differences seem to make an appreciable difference. The main feature of all these tests is the production of words. For this reason, the definition of this factor is given as the ability to produce words from a restricted area of meaning.

A more recent study in the Aptitudes Project (Guilford, Kettner, & Christensen, 1955) included all tests in the list above except Controlled Associations II. In that study these tests produced two factors, with Associations IV and Controlled Association leading one factor and with Associations III and Vocabulary Completion leading the other factor. The first factor was called *associational fluency* and the second one was not identified pending the outcome of the present analysis. It can now be suggested that the second one might have been *eduction of conceptual correlates,* the Vocabulary Completion test being in common with Factor M in this study.

Factor Q. Originality (O)

		A	B	C
4.	Cartoons	—	—	.47
32.	Plot Titles (cleverness)	—	—	.43
48.	Unusual Uses	—	—	.35 (.48 Pe, .31 SX)
8.	Consequences (remote)	—	—	.32 (.34 IF, .41 SX)
9.	Controlled Association	—	—	.30 (.56 AF)

This is the *originality* factor that was found previously in the creativity study where it was defined as the ability to produce remote, uncommon, or clever responses. These were the three sub-hypotheses for the *originality* factor. It was expected that this factor might be replaced by two or three factors if given the chance. Such apparently is not the case. The tests in the list above represent all three sub-hypotheses. Thus it seems that all three properties should be included in the definition of the factor.

Controlled Association has a significant loading on the factor but was not included on any of the sub-hypotheses for *originality.* It did have a barely significant loading on *originality,* however, in the creativity study, so it is not completely surprising that it has a significant loading here. The Quick Response test which was included under Sub-hypothesis 9*b* had a significant loading on *originality* in the creativity study but not here. Perhaps this is due to the fact that the "uncommonness" key used here was not derived from the sample that was tested.

Three objectively scored tests were also included among the tests for *originality*—Associations III, Associations IV, and Remote Verbal Similarities. None of these tests had significant loadings on the factor.

The two tests under Sub-hypothesis 9*c* (producing clever responses) have the highest loadings on this factor. It is possible that this factor should be defined as the ability to produce clever responses even though tests from the other sub-hypotheses have significant loadings here. It is possible to produce clever responses for the other tests that have significant loadings on the factor. If only the clever responses had been scored in these tests, they might have had higher loadings than they did. At any rate, it is quite possible that those individuals who produce clever responses to these tests produce a larger number of responses. This would explain why tests besides Plot Titles (cleverness) and Cartoons had significant loadings on the factor even though they do not obviously fit the sub-hypothesis "producing clever responses," and permit the defining of originality as "cleverness."

FACTOR R. REDEFINITION (RE) AND JUDGMENT (J)

	A	B	C
19. Gestalt Transformation	—	—	.48
25. Logical Classification	—	—	.44 (.39 V)
37. Seeing Deficiencies	—	—	.40 (.39 SP)
2. Associations III	—	—	.35 (.42 V, .36 AF)
36. Remote Verbal Similarities	—	+	.34

The two leading tests on this factor were among the tests hypothesized to contain *redefinition* variance. Both of these tests have had significant loadings on *redefinition* in previous studies. Two other tests, Remote Verbal Similarities and Seeing Deficiencies, have had significant loadings on the *judgment* factor in previous studies. As stated under the discussion of the hypotheses for *redefinition*, tests of both these factors have come out together on a single factor in several analyses. Since the two factors have not appeared separately in the same analysis, it is not known whether the two factors are distinct. No new light is thrown on the situation here except that a factor similar to the previous ones has reappeared.

Redefinition has been defined as the ability to redefine the uses of objects and the contents of statements. This definition does not fit Tests 36 and 37 very well. *Judgment* was defined in the second reasoning study as the ability to make wise choices in a somewhat ambiguous situation where it is necessary, in general, to make assumptions over and above the facts that are given. This definition fits all the tests in the list above except Associations III. From all the evidence, it appears that this is a confounding of two factors that were probably confounded once before.

FACTOR S. IDEATIONAL FLUENCY (IF)

	A	B	C
33. Plot Titles (low quality)	—	—	.60
7. Consequences (low quality)	—	—	.58
8. Consequences (remote)	—	—	.34 (.32 O, .41 SX)

The two leading tests on this factor were included in this study as reference tests for *ideational fluency*. This factor was defined in the creative-thinking study as the speed of calling up ideas in a situation in which there is relatively little restriction and quality does not matter. Both of the scores for the Consequences test have significant loadings. Evidently the low-quality responses were not completely separated from the remote responses in the scoring, as is indicated by the fact that the two scores have an intercorrelation of .39.

FACTOR T. SPONTANEOUS FLEXIBILITY (SX)

	A	B	C
8. Consequences (remote)	—	—	.41 (.32 O, .34 IF)
28. Object Synthesis	—	—	.40 (.40 NA)
48. Unusual Uses	—	—	.31 (.48 Pe, .35 O)

In the creativity study, *spontaneous flexibility* was defined as the ability to produce a diversity of ideas. In that study both Consequences (remote) and Unusual Uses had significant loadings. The other test on this factor, Object Synthesis, does not fit the definition very well since only one answer is required for each item. It may very well be, however, that having a diversity of ideas is helpful in arriving at an answer to the items in this test. At any rate, because of the similarity between this factor and the one found in the creativity study, it is called *spontaneous flexibility.*

FACTOR U. UNIDENTIFIED (U)

	A	B	C
39. Seeing Trends	.59	+	—
34. Prescribed Relations	.41	—	—

No attempt is made to interpret this doublet since the common element of these two tests is not obvious.

The remaining factors in the three batteries are residuals. None has a loading greater than .30 in absolute value.

DISCUSSION

Relations of the Factors to Those in Previous Studies

This study had several aims. One of these was the verification of factors found in previous studies. Eleven factors were under investigation. Five of these had been found in two or more studies in the Aptitudes Project. Four of the five appeared in this study in substantially the same form as before. These factors are: *eduction of perceptual relations, eduction of conceptual relations, eduction of correlates,* and *redefinition* (confounded with *judgment*).

Five of the six factors that had been found in only one study previously were verified. These factors are: *eduction of patterns, associational fluency, originality, sensitivity to problems,* and *verbal classification.*

Facility with verbal relations did not reappear. In the evaluation analysis, in which it had been previously found, the relatively large and somewhat heterogeneous list of tests related to it made interpretation difficult. Five of the tests related to it in the evaluation study and also present in this study proved to have significant loadings on *verbal comprehension, verbal classification,* and *associational fluency.* The first two of these appeared in the evaluation study. The inference is that the factor called "facility with verbal relations" was a composite of *associational fluency* plus other unidentified factors.

Tests for three reference factors—*verbal comprehension, general reasoning,* and *ideational fluency*—were included as marker tests. These three factors emerged as expected. In addition, the previously known factor of *spontaneous flexibility* emerged even though no marker tests were included for it.

Five new factors were discovered and named: *perceptual classification, eduction of structural relations, naming abstractions, eduction of conceptual correlates,* and *penetration.* There was some indication of another new factor that could be called "structural classification," but it was not clearly to be distinguished from the new correlates factor. Perhaps it is incorrect to say that the factor of *penetration* is new. Such a factor had been somewhat anticipated in the creativity study and a doublet factor could have been so identified.

Relation of the Factors to the Hypotheses

The discussion of this subject will follow fairly closely the order of discussion of the hypotheses given in the early part of this paper. First, some more general questions were asked pertaining to a few of the factors, then some specific hypotheses were stated pertaining to each factor under investigation.

The question of degree of generality was raised concerning two of the creative-thinking factors—*sensitivity to problems* and *originality.* In this study, *sensitivity to problems* appears to be somewhat more general than in the creativity study. It pertains to seeing defects or deficiencies in things other than mechanical gadgets and social institutions; more specifically, this extends to familiar, nonmechanical objects and to planned operations.

As to *originality,* on the surface it seems that the factor is as broadly general as it was before, pertaining to the production of unusual, remotely connected, or clever responses. The discovery of a *penetration* factor, however, suggests that remote associations may be dependent upon a factor other than *originality* or that there are different kinds of remoteness involved in tests of these two factors.

Questions were raised concerning the possible identification of two pairs of factors. The first pair included *associational fluency* and *facility with verbal relations.* It was found that the latter factor did not emerge, as such, in this study and that when previously found it probably represented a confounding of *associational fluency* with other unknown factors. The second pair was composed of *redefinition* and *judgment.* Tests for the two came out loaded on the same factor, but owing to the lack of traditionally definitive tests for *judgment,* the issue is still uncertain.

The general question was raised as to whether tests that call for classifying activity or the recognition of similarities among objects would go into the same categories as the tests that call for recognition of relations between objects. From the results of this study we have strong evidence that there is a group of factors representing abilities to classify objects separate from the group previously known representing abilities to cognize relations. There appear to be as many as three in each group, having to do with figural, structural, and verbal material, respectively.[8]

The introduction of a number of naming tests into the study raised the question of whether there would be one new factor representing the ability to produce names or verbal descriptions for ideas or perhaps several new verbalizing factors, or none. The results seem to be strongly on the side of the hypothesis of one such naming or verbalizing ability. It does not seem to matter whether the thing to be verbalized is a class or a relation or a system, so long as it is an abstraction of some kind.

The fate of the specific hypotheses concerning many of the factors may be treated very briefly:

Eduction of perceptual relations. The main hypothesis tested was whether seeing relationships and seeing classes among figural objects call for one primary ability or two. The results indicate two separate abilities, the other being called *perceptual classification.*

Eduction of conceptual relations. The hypotheses and findings parallel those for the preceding factor. There is a factor of *verbal classification,* found for the second time. Although there was no prediction of it at the time this study was initiated, there emerged a factor that could be identified as "structural classification," but since it was a doublet and bore apparent similarity to another structural factor, it was not accepted as a demonstrated factor.

Eduction of conceptual patterns. Tests were included to determine whether this ability extended to discovering rules and principles, also trends, as well as patterns. The answer seems to be that it pertains to all but seeing trends. Tests of the latter type go with tests of seeing relations. A trend may be regarded as a linear pattern or a succession of examples of the same relationship end to end, and hence lacks the complexity that is characteristic of tests of this factor. The results indicate another kind of conclusion, and that is the probable restriction of this ability to perceptual and structural tests. The qualifying term "conceptual" has therefore been dropped from the name of the factor. On the basis of other distinctions being found between tests of figural and structural materials, we may naturally ask the question whether there should not be two factors in place of the one obtained, and perhaps a third pertaining to conceptual materials.[9]

Eduction of correlates. The main hypothesis tested was whether there would be two factors—perceptual and conceptual—to replace the one that had been previously found. The results strongly favor the distinction between two, the one being conceptual, but the other being heavily featured by structural (letter) material. This raises the question of the existence of a third factor of eduction of correlates among figures.

Symbol manipulation. The answer pertaining to this factor is that there are two distinct primary abilities, the other being designated as *symbol substitution.*

Verbalizing ability. The result concerning this hypothesis turned out to be one new factor of abstraction naming, which pertains to various kinds of abstractions, as indicated in the general discussion.

Sensitivity to problems. This factor seems to be a little broader than when first discovered, as explained in the general discussion.

Associational fluency. The new, positive information concerning this factor indicates that it can be tested by calling for few or many responses to stimulus words. The finding of a new factor of *eduction of conceptual correlates,* however, which also involves the giving of a meaningful response that satisfies a given relationship, calls for some kind of discrimination between the two. The best basis at present is that the latter factor involves one best or most conventional response whereas *associational fluency* involves a variety of responses. This kind of discrimination contributed to the distinction between "convergent" and "divergent" productive thinking, which played an important role in the development of the "structure of intellect" (Guilford: 1956, 1957).

---

[8] It was the results such as these in the present investigation that gave the strongest impetus to attempts to organize the intellectual factors in a system that has been called the "structure of intellect" (Guilford: 1956, 1957, 1958).

[9] Subsequent to this study, it has been proposed that the well-known factor of *spatial orientation* is the parallel, figural *eduction of patterns,* and that *general reasoning* is the parallel, conceptual *eduction of patterns* (Guilford: 1956, 1957).

Originality. Although, as indicated above, *originality* still seemed to be involved in tests calling for uncommon, remotely associated, and clever responses, there is some indication that remote associations may be also an indicator of the new factor of *penetration.*

Redefinition. The separability of this factor from that of *judgment* is still an open question.

Facility with verbal relations. The indications are that this factor, which appeared only once before, was then a composite. It did not reappear, as such.

Verbal classification. The ability to recognize classes of things presented verbally was confirmed, with the factor supporting the hypothesis of "seeing conceptual similarities."

Categories of Factors

Earlier work in the project, and this particular study, recognized the general domains or categories of reasoning, creativity, planning, and evaluation. While these categories have had heuristic value, it has become rather clear from the results of the present investigation that some new categories are in order. The category of evaluation is the only one that seems to have persistent value.

There has been increasing recognition that certain factors represent abilities to discover information, or to rediscover or recognize it, in the form of things, relations between things, classes, patterns or systems, and so on. Other factors have been distinguished from discovery factors by the fact that they involve production of new information from given information. They can be called production factors. Within this group a further distinction has developed, as suggested above: that between convergent thinking and divergent thinking. The former is aimed toward a conventional response of some kind while the latter seeks a variety of responses.

It is possible to give the former, heuristic categories more empirical meaning by reference to these new factor-category concepts. Reasoning may be defined as relational thinking. We cognize relations, we use relations in producing new information, and we evaluate the products of these steps in terms of logical consistency. Creative thinking seems to depend most directly upon the divergent-thinking category of abilities. In a broader sense, creative production, as in ordinary problem solving, also clearly involves other than divergent-thinking operations.

Improvement of Tests

The third objective of this study was more incidental than the others. Only general implications will be given here concerning the kinds of improvements that are indicated for tests as a consequence of new information. In addition to knowledge of the statistical data concerning each test we also learn something concerning its factorial validity; what common factors it measures in the population involved and to what extent. We thus learn which tests are best for each factor and what secondary factor variances to attempt to eliminate in future forms of those tests. We learn more about the properties of some of the factors so that perhaps new and better types of tests can be constructed. The results of this investigation carried its expected share of such implications.

SUMMARY

The main purpose of this study was to investigate in the same analysis certain factors that had been found previously in the domains of reasoning, creativity, and evaluation. Three major objectives were: (*a*) verification of factors found only once or twice before, (*b*) clarification as to their scope and properties, and (*c*) derivation of information that would lead to improvement of tests measuring these factors.

Eleven factors were selected for investigation, excluding those of doubtful nature and those needing more intensive study than could be given them in such a general survey study as this one. General questions asked pertained to the possible identity of factors found in different studies, to the generality or scope of certain factors, and to the possibility that some formerly apparently unitary factors might be replaced each by two. The introduction of a number of tests calling for the examinee to name or describe relations, classes, or patterns called for hypotheses regarding the possible finding of one or more new verbalizing factors. For most expected factors alternative hypotheses were stated as to their properties.

Fifty-seven tests were selected, adapted, or constructed to test the hypotheses, with some of them being introduced as marker tests of reference factors (other than the 11 being investigated). The twelve-hour battery had to be broken up into three partially overlapping batteries. Each battery was administered to a sample of approximately 200 air cadets, two in the Air Force and one in the Navy. Three factor analyses were done by the centroid method followed by orthogonal rotations of axes.

Twenty psychological factors were interpreted, five of which were reference factors. Nine of the 11 factors under investigation were found in somewhat the same form as before, with some clarification as to their unique properties. They are: *eduction of perceptual relations, eduction of conceptual relations, verbal classification, eduction of patterns, eduction of correlates, sensitivity to problems, associational fluency, originality,* and *redefinition* (probably confounded with *judgment*). *Facility with verbal relations* did not reappear, as such, hence it was probably previously a composite factor. *Symbol manipulation* appeared but some of the symbolic tests determined another factor, distinct from it—*symbol substitution.*

Five new factors were found: *perceptual classification* (ability to recognize classes of figural objects), *eduction of structural relations* (ability to see relations between structural objects, such as groups of letters), *naming abstractions* (ability to name or describe abstractions such as classes, relations, and patterns), *eduction of conceptual correlates* (ability to supply a response when needed to complete a meaningful whole), and *penetration* (ability to see remote connections or similarities). A doublet factor might have been interpreted as "structural classification" (an ability to recognize classes of structural objects), but it could not be clearly distinguished from *eduction of structural relations.*[10]

A consideration of natural groupings of factors suggests that the heuristic categories of reasoning and creative thinking would be more appropriately replaced by categories of discovery, or cognition, and of productive thinking. The latter could be subdivided into a convergent-thinking and a divergent-thinking group. The category of evaluation seems to remain useful.

---

[10] In current writings the term "structural" has been replaced by "symbolic" and the term "conceptual" by "semantic" in order to reduce ambiguities.

REFERENCES

ADKINS, D. C., & LYERLY, S. B. *Factor analysis of reasoning tests.* Chapel Hill, N. C.: Univer. North Carolina, 1951.

CARROLL, J. B. A factor analysis of verbal abilities. *Psychometrika,* 1941, **6,** 279-307.

GREEN, R. F., GUILFORD, J. P., CHRISTENSEN, P. R., & COMREY, A. L. A factor-analytic study of reasoning abilities. *Psychometrika,* 1953, **18,** 135-160.

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.,* 1956, **53,** 267-293.

GUILFORD, J. P. A revised structure of intellect. *Rep. psychol. Lab.,* No. 19. Los Angeles: Univer. Southern California, 1957.

GUILFORD, J. P. New frontiers of testing in the discovery and development of human talent. In *Seventh annual western regional conference on testing problems.* Los Angeles: Educational Testing Service, 1958.

GUILFORD, J. P., & CHRISTENSEN, P. R. A factor-analytic study of verbal fluency. *Rep. psychol. Lab.,* No. 17. Los Angeles: Univer. Southern California, 1956.

GUILFORD, J. P., CHRISTENSEN, P. R., KETTNER, N. W., GREEN, R. F., & HERTZKA, A. F. A factor-analytic study of the Navy reasoning tests with the Air Force Aircrew Classification Battery. *Educ. psychol. Measmt.,* 1954, **14,** 301-325.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation: I. Hypotheses and descriptions of tests. *Rep. psychol. Lab.,* No. 11. Los Angeles: Univer. Southern California, 1954.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. The relation of certain thinking factors to training criteria in the U. S. Coast Guard Academy. *Rep. psychol. Lab.,* No. 13. Los Angeles: Univer. Southern California, 1955.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation: II. Administration of tests and analysis of results. *Rep. psychol. Lab.,* No. 16. Los Angeles: Univer. Southern California, 1956.

GUILFORD, J. P., & LACEY, J. I. (Eds.) Printed classification tests. *Army Air Forces Aviation Psychol. Program Res. Rep.,* No. 5. Washington, D. C.: Government Printing Office, 1947.

HERTZKA, A. F., GUILFORD, J. P., CHRISTENSEN, P. R., & BERGER, R. M. A factor-analytic study of evaluative abilities. *Educ. psychol. Measmt.,* 1954, **14,** 581-597.

KETTNER, N. W., GUILFORD, J. P., & CHRISTENSEN, P. R. A factor-analytic investigation of the factor called general reasoning. *Educ. psychol. Measmt.,* 1956, **16,** 438-453.

WILSON, R. C., GUILFORD, J. P., CHRISTENSEN, P. R., & LEWIS, D. J. A factor-analytic study of creative-thinking abilities. *Psychometrika,* 1954, **19,** 297-311.

ZIMMERMAN, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika,* 1946, **11,** 51-55.

(Accepted for publication July 29, 1958)

Vol. 73, No. 10 Whole No. 480, 1959

Psychological Monographs: General and Applied

GROUP PSYCHOTHERAPY AND CONCOMITANT PSYCHOLOGICAL CHANGE[1]

EARL J. ENDS[2] AND CURTIS W. PAGE[3]

Willmar State Hospital, Minnesota

THE DATA and conclusions herein presented represent an interim report covering the second phase of a continuous research program the over-all purpose of which is to assess the usefulness of group psychotherapy with alcoholic patients and to determine the most efficient type and intensity of application. Further goals are to seek clearer understanding of the therapeutic process itself for purposes of training therapists, selecting patients, discovering and applying whatever modifications of the therapeutic process may be necessary to widen the applicability of psychotherapy to an inebriate hospital population. The ultimate practical goal is as always the improvement of the recovery rate.

In reading this report and in evaluating our results and conclusions, it is important that the reader have some general idea of the total treatment program and the generally involuntary nature of hospitalization for the treatment of alcoholism because both of these factors materially influence the interpretation of our data.

The formal aspects of the alcoholic program at Willmar State Hospital include a series of 26-hour-long orientation lectures covering psychological, physical, social, and spiritual problems associated with alcoholism; a thorough grounding in Alcoholics Anonymous (AA) history, philosophy, and principles; several lectures on personality growth and development and on mental mechanisms of a pathological nature. This part of the program is primarily educational in scope and purpose, but in effect is often also inspirational. These lectures, given by 13 different staff members, serve as the core of the general treatment program. The purely medical problems associated with extremely excessive drinking and alcoholism are treated whenever they appear with a thoroughness and degree of skill perhaps unmatched anywhere.

In addition to the orientation lecture series, numerous more specific lecture and discussion groups are part of the regular program. The AA program is considered to be an integral part of the general treatment program. Group AA meetings are held four nights a week. Visiting AA speakers and/or selected speakers are present twice a week, the hospital chaplain conducts one meeting weekly on the spiritual aspects of AA, and one meeting is devoted to instructions on the "12 steps of AA." A group question-and-answer hour is held weekly by one of the alcoholic counselors. The Toastmasters' Club meets once weekly offering an opportunity for alcoholics to master the art of public speaking. In addition, the alcoholic is invited to talk his problems over with one of the four AA counselors on the staff. AA social discussion groups, comprising eight to ten alcoholics, meet three times weekly with the AA counselors and the chaplain. Although AA participation is voluntary, the majority of alcoholics do participate in some phase of the AA program. Group psychotherapy is offered only as staff time permits. There are usually at least two groups in progress aside from those used in research projects at any time except during the summer months when staff vacations and short terms of hospitalization make group therapy rather impossible to maintain.

The Willmar program is currently open to both voluntary and committed alcoholics. Commitment as an alcoholic, of course, involves legal proceedings and a court hearing. Voluntary admissions need only come to the hospital seeking help for alcoholism and agree to stay the specified period. The treatment program is scheduled for 60 days and, after admission, no distinction is made between voluntary and committed alcoholics. Currently about half of the admissions are voluntary; however, these latter are not what they seem. It is probable that less than 25% of voluntary admissions truly come of their own free will seeking help for a self-recognized problem. The other 75% or more "volunteer" under coercion of one kind or another. Often they have been threatened with commitment if they refuse to volunteer, others have been threatened with loss of their jobs, divorce, jail terms, and many other things of a like nature.

---

[1] Supported in part by a grant from Department of Public Welfare, State of Minnesota. We are indebted to Nelson J. Bradley, Superintendent, for early and continuing administrative support of this research program, to M. Duane Sommerness for criticisms and encouragement at the inception of the program, to the department of psychology, and specifically Jean J. Rossi, for an indispensable role in carrying out the program, to Lloyd A. Smith and other physicians, ward personnel, and hospital personnel for their cooperation.

[2] Presently with Psychological Research and Development Corp., Fullerton, California.

[3] Presently Director of Research and Training, Traverse City State Hospital, Michigan.

The over-all effect of these practices as far as psychotherapeutic treatment is concerned is that we are *not* dealing with people who are motivated for therapy. Many of them are extremely resentful over being here and are hostile and bitter toward those responsible. Arbitrary assignment to group therapy under the demands of the research project results, usually, in even more resentment and hostility directed toward the hospital staff in general and the therapist in particular. In terms of usual psychotherapeutic thinking, then, therapy begins under extremely unfavorable circumstances with *S*s who at best are considered very poor therapeutic risks.

Reprints of the interim report (Ends & Page, 1957) on the initial phase of the project are available. In this first phase we sought to determine the most effective of three theoretical approaches to group psychotherapy using the analytic, the group-centered (Rogerian), and a learning theory approach based on the work of Mowrer (1950, 1953). A nontherapy discussion group was included as a control. Conclusions based primarily on a complex analysis of *Q*-sort data were that the Rogerian group-centered approach possessed the widest applicability and greatest effectiveness, all things considered. Follow-up data supported our latter conclusion.

Unfortunately, overcoming some of the shortcomings of the first phase of the study necessarily results in a loss of precision in comparisons of results of the two phases, but enough similarity remains so that such comparisons are both legitimate and revealing.

In addition to providing a test of several new hypotheses, the second phase was also designed to serve as a check on our earlier conclusions regarding the effectiveness and *modus operandi* of the group-centered approach to psychotherapy. The control group used for the second phase received unrestrainedly all the benefits of the general treatment program, including group meetings with AA counselors, individual counseling by chaplains, individual counseling by AA counselors, and individual counseling by psychologists just as did the group therapy *S*s. However, no *S*s in either therapy or control groups received regularly scheduled intensive individual psychotherapy by psychologists or psychiatrists. As far as treatment in the second phase was concerned, the only difference was that the experimental group of inebriates received group therapy in addition to the regular treatment program whereas the control group did not. It is important to keep in mind that improvement and change in the control group represents the effect of the general treatment program, *not* "spontaneous" change in the absence of treatment.

Some of the hypotheses which the second phase was designed to test or explore depend on follow-up data to be acquired at specified time intervals after discharge from the hospital. Since the data will not be collected and ready for analysis for some time, reference to those hypotheses will be omitted in this interim report.

PROBLEM

Does Rogerian group-centered therapy, when added to the general alcoholic treatment program, result in therapeutic movement quantitatively and/or qualitatively different from the general alcoholic treatment program alone? Secondly, does therapeutic movement as evident in data from the first phase of the project continue and progress to a significantly greater extent when the number of therapy sessions is doubled from 15 sessions in the first phase to 30 sessions in the second (current) phase? A third problem concerns the nature and rate of change of therapeutic movement. Do increasing amounts of therapy result in increasing, diminishing, or proportional returns in therapeutic progress?

MAJOR HYPOTHESES

1. The addition of Rogerian group-centered therapy to the general treatment program will result in both qualitative and quantitative differences in therapeutic change.

2. Doubling the number of therapeutic sessions from 15 to 30 will result in significantly greater therapeutic movement.

3. Doubling the number of therapeutic sessions will result only in quantitative rather than qualitative differences, i.e., the pattern of change will be the same for both the 15-session and the 30-session therapy groups.

4. As the number of therapy sessions increases, the rate of therapeutic progress will increase. Other minor hypotheses explored concerning the therapeutic process in particular will be stated at appropriate points in the discussion.

METHODS AND PROCEDURES

Selection of S*s*

The alcoholic population from which *S*s were selected for participation in the second phase of the study was defined by the following criteria: male, white, within the age range of 21 to 60, inclusive, without severe organic pathology (CNS) or psychotic impairment according to available psychiatric opinion, and above the forty-fifth percentile on the Army General Classification Test (AGCT). No distinction was made between committed and voluntary admissions.

Since admissions are continuous, 7 days a week, discharge dates (60 days later) are also continuous. It was, therefore, necessary to select experimental and control groups from a series of 8-day admission periods in order that *S*s could be tested, groups organized, groups given six weeks of group therapy or control conditions, and given the post-treatment testing before the 60-day limit had expired for any of the group members. This means that no *S* was hospitalized less than 5 days or more than 12 days before his initial test which was always the AGCT.

All those who met the age requirement, who were judged neither psychotic nor organically impaired by the ward psychiatrist, were not confined to bed for emergency medical treatment, and who were admitted for treatment for alcoholism during any of the 8-day periods used, were then given the AGCT. Those achieving below the 45 percentile were given no further tests. The rest were given the Minnesota Multiphasic Personality Inventory (MMPI). Those having MMPI profiles that clearly suggested psychosis were eliminated at this point. The remainder of the group, usually about 15 or 16 out of the initial 30 admissions, was reduced to 14 by disqualifying those with the most psychotic-like MMPI profiles. These were then given self and ideal-self *Q* sorts. From this eligible group, seven were randomly assigned to a therapy group, the remainder to a control group. Because only two experienced therapists were available, and each could carry only one group at a time in addition to other duties, selection was not a weekly process but occurred every third week for the groups used in this project. Selection for all other psychotherapy groups of alcoholics for training and treatment purposes follows the same general procedure, but such other groups are not part of this study and need not concern us here.

Method

All *S*s in therapy and control groups were given pre- and post-MMPI's and self and ideal-self *Q* sorts immediately preceding and immediately following the six-week therapy or control period. In addition, the therapy groups were also given self and ideal-self *Q* sorts at the end of the 10th and 20th therapy sessions.

All therapy sessions were tape recorded to permit continual spot checks on therapeutic technique by the psychology department staff members to insure that the theoretical orientation (Rogerian group centered) was not seriously violated by either therapist.

Each of the two therapists carried three of the experimental (therapy) groups. Each group began with seven inebriates. However, the first four groups finished five men each and the last two groups finished only four men each. Attrition was due to discharge from the hospital for various reasons.

With the exception of daily therapy meetings and the two extra *Q*-sorting sessions (at the end of the 10th and the 20th therapy sessions) for those in the therapy groups, control and experimental *S*s were subjected to the same routine treatment. Indeed, with the further exception of the group testing involved, experimental and control *S*s received the same treatment as the rest of the alcoholic population, a necessary condition for the testing of our first hypothesis.

For purposes of analysis, all therapy groups were lumped together and will hereafter be referred to simply as the *therapy group* which totals 28 men. The control group was similarly consolidated, but since the attrition for controls was slightly higher, one extra group of controls was selected, following identical procedures, to bring the control group up to a comparable number. Of this extra group, four *S*s completed the control period, but only three were wanted and were chosen by use of a table of random numbers. Thus both groups contained 28 men each.

In order to provide some estimate of the reliability or stability of the self and the ideal-self concepts and other relevant measures, a third series of groups was selected in an identical manner using the same criteria. These groups, hereafter referred to collectively as the *two-week control group,* were given a second self and ideal-self *Q* sort and a second MMPI exactly two weeks after their initial tests. This is admittedly a second-best way of obtaining "reliability" estimates, but was necessitated by several factors. First, this estimate could have been obtained from the therapy group only by sacrificing two weeks of therapy; second, it could have been obtained from the control group at the risk of distortion resulting from a possible reaction of being excluded from the therapy groups, a possible effect that we assume does not influence final testing of the control group because of the comparatively long period of elapsed time (six weeks). This factor could not influence results on the two-week control group because selection was postponed until all therapy groups were terminated. Actually, since all alcoholics are subjected to systematic treatment pressures almost from the moment they are admitted for treatment, true reliability of measures or concepts is indeterminate for our population. The statistics which we will report as estimates of reliability derived from tests given two weeks apart to the two-week control group provide an estimate of the relative stability of the measures over the stated period in *S*s who are under a systematic program designed specifically to produce favorable change in their behavior and in their general personality functioning. However, these statistics *do* provide us with measures to which we can compare similar figures for the therapy group following two weeks of therapy.

Subjects

Age, AGCT scores, years of education, current marital status, number of children, total years of drinking, and number of years of uncontrolled drinking for the various groups used in the study are presented in Table 1.

Some differences in the groups on the variables listed in Table 1 are readily apparent. However, with the exception of the 15-session therapy group which for the first phase of the project was drawn from the 25-45 age range of the defined population, the groups depart from randomness only by virtue of undefined biases in those portions of our initial samples which were lost due to early discharge. Attrition was highest for the full-term (6-week) control group. That any of the listed variables are more than indirectly relevant is doubtful.

Method of Analysis

Q *technique.* The method of analysis of *Q*-sort data developed to extract differential patterns of change for the first phase of the project was retained. As briefly as possible, the procedure is outlined in the next few paragraphs.

Individual product-moment correlation coefficients were calculated on each *S* for each of the 10 combinations of pre-self, pre-ideal, post-self, post-ideal, and criterion-sort, taken two at a time. *Pre* and *post* refer to both therapy and specified control periods.

The concepts of self and ideal-self (hereafter referred to simply as *ideal*) are admirably defined by Hartley (1951), Rogers (1942, 1951), and Rogers and Dymond (1954) and need not be repeated here. They are operationally limited by the particular set of cards used. To avoid the necessity of continually defining words during testing, the phrasing of some of the more difficult items in the sort developed by Butler and Haigh (see Rogers & Dymond, 1954) and used extensively by the Chicago group, was transposed into more familiar language for use in the second phase of our project (see Appendix).

The criterion-sort (C-S) was developed for use in this study to provide some independent and more or less objective measure of the direction of movement of the self or ideal toward or away from mentally healthy configurations (see Appendix). The C-S is the mean pattern reflecting the highest degree of healthy psychological adjustment possible with the particular sample of statements used according to the statistically combined *Q* sorts of five experienced clinical psychologists performed under instructions to sort the cards so that in his opinion the distribution would represent the healthiest possible self concept achievable with the particular set of cards used. Correlations between the psychologists' sorts so performed and the derived criterion sort ranged from .87 to .92.

All correlation coefficients were converted to Fisher's (1954) z's throughout. Since we are interested in magnitude and direction of movement rather than in absolute size of z's in a study of therapeutic change, differences between *pairs* of z's give us the most information. This is because they reflect change. For example, by subtracting the pre-self:post-self z from the pre-ideal:post-ideal z, it is possible to determine whether ideal or self concept changed more or, conversely, which was more stable. The difference scores so obtained for each *S* were summed algebraically for each group for each of the comparisons made. Group mean differences were then calculated for the analysis of change for each group. These mean differences are hereafter referred to as *indexes* for simplicity. Error terms for estimating the significance of such differences between correlated group means, i.e., those based on the same group of *S*s, were calculated using McNemar's (1949, p. 226) Formula 92; significance of differences between uncorrelated group mean z's were calculated using Formula 91 (McNemar, 1949, p. 224).

The following indexes all seem to yield unique information:

1. G minus D:[4] Tells whether the self tended to grow healthier or less healthy over the therapy or control period. Plus means adjustment increased.

---

[4] The code letters refer to the two correlations used in the derivation of the indexes and are included here for reference. They are explained further in the next section and the respective sort correlations which they represent appear in Table 2.

TABLE 1

SUBJECTS WHO COMPLETED THERAPY OR CONTROL PERIODS

		Groups			
		Therapy 30s (6 weeks) *N*=28	Control (6 weeks) *N*=28	Control (2 weeks) *N*=28	Therapy[a] 15s (6 weeks) *N*=15
Age	Mean	45.107	42.821	44.464	36.273
	SD	5.725	9.294	7.826	5.804
AGCT	Mean	115.857	116.607	120.357	112.600
	SD	10.242	13.316	11.650	10.294
Years education	Mean	11.107	9.642	11.071	9.733
	SD	2.610	3.096	2.975	1.526
Marital status	Married	39.3%	53.6%	42.8%	53.3%
	Single	28.6%	17.8%	14.3%	26.7%
	Separated	7.1%	7.1%	14.3%	6.7%
	Divorced	14.3%	17.9%	25.0%	13.3%
	Widowed	10.7%	3.6%	3.6%	0.0%
No. of children[b]	Mean	3.200	2.347	1.857	1.800
	SD	2.731	1.879	1.761	1.720
Total years drinking	Mean	24.178	21.643	23.857	18.600
	SD	6.717	7.779	7.791	8.333
Years uncontrolled drinking	Mean	10.214	6.185	7.214	2.800
	SD	9.060	4.850	6.966	2.638

[a] This group is from the first phase of the project (Ends & Page, 1957).

[b] Includes step-children; single men are omitted from this statistic.

2. F minus B: Tells whether the self and ideal converged or diverged during the course of the therapy or control period. Plus means concepts converged.

3. H minus A: Tells whether ideal or self changed more. Plus means the ideal was more stable.

4. J minus I: Tells whether the ideal tended to grow healthier or unhealthier during the course of the therapy or control period. Plus means ideal became healthier.

5. E minus B: Tells whether the self moved toward or away from the pre-ideal during the course of the treatment or control period. Plus means self moved toward pre-ideal.

6. F minus C: Tells whether the self moved toward or away from the post-ideal during the course of the therapy or control period. Plus means self moved toward the post-ideal.

7. C minus B:[5] Tells whether the ideal tended to move toward or away from the pre-self during the therapy or control period. Plus means ideal moved toward the pre-self.

8. F minus E: Tells whether the ideal moved toward or away from the post-self during the therapy or control period. Plus means the ideal moved toward the post-self.

Analysis of MMPI protocols for the various groups consists simply of comparisons of mean *T* scores for the *L, F, K, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma,* and the raw scores for the Barron (1953) Ego Strength Scale (*Es*). As with the

---

[5] The direction of subtraction for this index was reversed from that used in the previous report (Ends & Page, 1957) at the suggestion of Carl Rogers in a personal communication (April 15, 1957).

Q-sort correlations, significance of differences among means of the various groups both before and after therapy were calculated to round out the picture and provide a measure of the initial comparability of the groups, but since we are primarily concerned with discovering systematic differences in change on the various scales between groups, the emphasis is placed on analysis of pre-post difference scores.

RESULTS

Hypothesis 1

To reduce consideration of results to manageable proportions, complete statistical presentation will be reserved for the final report. Results relevant to specific hypotheses will be presented in abbreviated form under the heading of the given hypothesis.

Hypothesis 1: The addition of Rogerian group-centered therapy to the general treatment program will result in both qualitative and quantitative differences in therapeutic change to a significant extent.

Results of Q Sort

The mean z's for the 30-session (6-week) therapy group and for the 6-week control group for the 10 correlations determined for each *S* are presented in Table 2 along with the code designation for that particular correlation. For example, the pre-self: post-self correlation is coded *A*; the figures in the body of the table represent the means of all such correlations (in Fisher's z's) for the several groups.

Following the previously outlined procedure, subtracting a given mean z from another mean z from the same group, the indexes of movement were derived. These mean differences are given in Table 3 along with the level of statistical significance of the movement. The two capital letters at the head of each column refer to the two mean z's used in deriving the index. For example, Index 1 is based on G minus D or, decoding, is obtained by subtracting the pre-self:C-S mean z (D) from the post-self:C-S mean z (G) for the same group. The dash between the letters, e.g., H − A, is a minus sign indicating the direction of subtraction.

Therapy group. This group demonstrated significant movement on seven of the eight indexes. The positive movement on Index 1 indicates that this group achieved a much better adjusted perception of the self; Index 2 reveals that it gained enormously in self-acceptance by reducing the discrepancy between self and ideal over the therapy period. Index 3 indicates that the self changed more than did the ideal while Index 4 shows that the ideal grew significantly better adjusted during the therapy period. Index 5 reveals that the change in the self-percept over the therapy period was congruent with and perhaps influenced by

TABLE 2

MEAN z'S FOR THERAPY (30S) AND CONTROL GROUPS

Code	Sorts Correlated	Therapy (N=28)	*SD*	Control (N=28)	*SD*	SE$_{MD}$	*d*	p[a]
A	Pre S—Post S	.562	.225	.494	.232	.062	.068	N.S.
B	Pre S—Pre I	.303	.299	.406	.320	.084	−.103	N.S.
C	Pre S—Post I	.349	.293	.310	.210	.069	.039	N.S.
D	Pre S—C-S	.349	.263	.361	.226	.067	−.012	N.S.
E	Post S—Pre I	.434	.314	.399	.266	.079	.035	N.S.
F	Post S—Post I	.702	.519	.502	.313	.116	.200	<.10
G	Post S—C-S	.507	.265	.475	.248	.070	.032	N.S.
H	Pre I—Post I	.695	.263	.549	.275	.073	.146	<.05
I	Pre I—C-S	.654	.243	.608	.242	.066	.046	N.S.
J	Post I—C-S	.738	.234	.650	.235	.064	.088	N.S.

[a] Two-tailed test of significance applied.

TABLE 3

Q-Sort Movement Index Means for Therapy Group (30s) and Control Group (6 Weeks)

	1 G−D	2 F−B	3 H−A	4 J−I	5 E−B	6 F−C	7 C−B	8 F−E
Group Therapy (N=28)								
M_D	.158	.399	.133	.084	.131	.353	+.046	.268
p^a	<.01	<.001	<.01	<.05	<.01	<.001	N.S.	<.001
SD_D	.264	.470	.222	.198	.243	.406	.175	.334
Control Group (N=28)								
M_D	.114	.096	.055	.042	−.007	.192	−.096	.103
p^a	<.05	N.S.	N.S.	N.S.	N.S.	<.001	<.10	<.02
SD_D	.288	.338	.286	.186	.311	.245	.262	.214
Movement Differences Between Groups								
M_D	.044	.303	.078	.042	.138	.161	+.142	.165
p^a	N.S.	<.007	N.S.	N.S.	<.07	<.08	<.02	<.03
SE_{MD}	.075	.111	.070	.052	.076	.090	.061	.076

[a] Probability for two-tailed test.

the relatively more stable pre-therapy ideal. However, Index 6 indicates that movement of the self during therapy was even more closely oriented toward the post-therapy ideal. In other words, the better adjusted, post-therapy self-percept appears to have been derived through a process of achieving previously held "ideals" as well as achieving new goals. This is strongly suggestive of a well regulated "growth" process. Index 7, the only index which yielded no movement significant (at least at the 10% level) for this group, indicates that the better adjusted post-therapy ideal was in closer congruence with the pre-therapy self than was the pre-therapy ideal. This movement, while not significant, at least suggests that modification of the ideal was not based on rejection of the pre-therapy self, nor was it based on a defensive affirmation of the pre-therapy self. It hints of greater acceptance of previously rejected self-percepts. Index 8 reveals that the post-therapy ideal had moved clearly and strongly toward the post-therapy self. When this movement is considered in conjunction with movement on Index 6, it becomes apparent that the two represent integrated movement of both self and ideal resulting in much greater self-acceptance. When considered simultaneously, these movements present a picture of healthy, well-integrated therapeutic growth over the therapy period.

Control group. Continuing with reference to Table 3, this group demonstrated significant movement on four of the eight indexes. The positive movement on Index 1 indicates that this group also achieved a better adjusted perception of the self, but the gain was less marked than for the therapy group. Index 2 reveals only a moderate, statistically insignificant gain in self-acceptance. Index 3 indicates that change in the self tended to be greater than the change in the ideal over the six-week control period. Here, too, the ideal showed a tendency to grow healthier during hospitalization as shown by Index 4. Index 5 indicates that the change in self over the control period was relatively independent of the pre-treatment ideal. However, Index 6 shows that the movement of self over the control period was closely oriented toward the post-treatment ideal. Differently put, the more adjustive post-treatment self appears to have been derived by a process of achieving newly held "ideals" or goals without making any significant headway toward previously held but unattainable self-concept goals. Index 7 indicates that the better adjusted post-treatment ideal was less congruent with the pre-treatment self than was the pre-treatment ideal. This suggests that modification of the ideal may have been based to some extent on a rejection of the pre-treatment self. Index 8 shows that the post-treatment ideal had moved toward the post-treatment self. When this movement is considered in conjunction with movement on Index 6, it is apparent that the two represent only mild integrated or coordinated movement of self and ideal, and result in a minimal, statistically insignificant increase in self-acceptance. Considered simultaneously, these movements present a picture of some improvement yet do not suggest a true growth process at all in terms of personality change as measured by the variables used. The relationships between movements do not seem to reflect a trend toward a more stable personality configuration; rather they seem to reflect change, usually in the "right" direction when variables are considered individually, but not a sound growth process which necessarily implies integrated change of relevant variables resulting in a more stable organization of concepts about the self, more stable because change has resulted from a unified reorganization process. The change shown by the control group, then, seems likely to be a temporary one because it does not appear to be supported by the careful thinking-through (such as occurs in therapy) necessary to stabilize it. A summary of the changes in the therapy and control groups is presented in Fig. 1.

MMPI Results

The therapy and control groups showed no differences between means on any of the MMPI scales including *Es* on pretesting that were significant at the 10% level or beyond. We will concern ourselves here with changes on MMPI scales between pre- and posttesting for the two groups. Table 4 presents data relevant at this point.

Therapy group. The therapy group demonstrated significant changes between pre- and post-means on several MMPI scales. The first of these was a decrease in the *F* score ($p = .02$); the second was an increase in the *K* score ($p = .10$). A sizable decrease occurred on the *D* score ($p = .01$) and an even larger decrease occurred on the *Pa* scale ($p = .001$). The *Pa* scale was the only one on which the therapy group scored a statistically significant greater change than the control group ($p = .05$). The therapy group also showed a significant decrease in *Pt* ($p = .02$) and a significant gain in ego strength (*Es*) ($p = .01$). Using literal interpretations, these changes may be summed up as a decrease in the psychological confusion reflected in *F*, an increase in psychological defensiveness or conversely a decrease in self-criticism as reflected in *K*, an increase in morale and optimism or conversely a decrease in depressive attitudes as reflected in *D*, a marked decrease in suspiciousness of the motives of others as indicated by *Pa*, a decrease in neurotic anxiety symptoms as reflected in *Pt*, and a gain in ego strength as reflected in *Es*. At the same time, there were statistically insignificant decreases on all other MMPI scales with the exception of *Ma* which showed an infinitesimal increase.

Control group. Continuing with reference to Table 4, this group also demonstrated

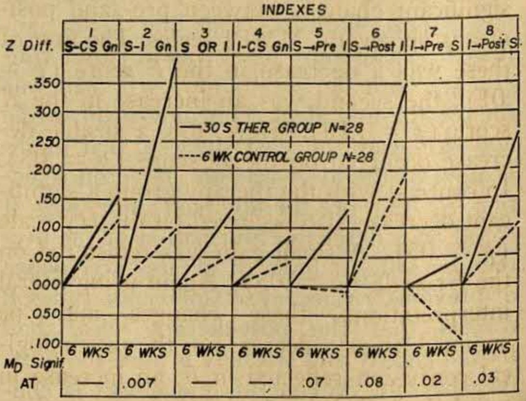


FIG. 1. Comparison of change in 30s therapy group and six-week control group.

TABLE 4

MEANS AND MEAN DIFFERENCES ON MMPI SCALES FOR THERAPY GROUP (30s) AND CONTROL GROUP (6 WEEKS)

Scale	Group ($N=28$)	Pre-Mean	Post-Mean	M_D	p
L	Therapy	48.607	48.107	.500	
	Control	48.928	48.876	.142	
F	Therapy	56.821	53.429	3.392	<.02
	Control	55.500	52.786	2.714	<.01
K	Therapy	51.357	54.179	−2.822	<.10
	Control	50.678	52.536	−1.858	<.10
Hs	Therapy	55.035	52.892	2.143	
	Control	52.214	51.285	.929	
D	Therapy	63.821	58.785	5.036	<.01
	Control	61.857	57.464	4.393	<.05
Hy	Therapy	57.857	56.964	.893	
	Control	56.500	53.928	2.572	<.05
Pd	Therapy	69.571	68.071	1.500	
	Control	65.285	65.286	.001	
Mf	Therapy	57.321	56.571	.750	
	Control	54.286	52.357	1.929	
Pa	Therapy	61.250	55.250	6.000	<.001
	Control	59.392	57.571	1.821	
Pt	Therapy	61.607	57.000	4.607	<.02
	Control	58.214	56.893	1.321	
Sc	Therapy	54.750	52.107	2.643	
	Control	51.642	51.714	−.072	
Ma	Therapy	57.678	57.750	−.072	
	Control	56.928	59.679	−2.751	
Es	Therapy	41.392	44.250	2.858	<.01
	Control	42.714	44.321	1.607	<.05

significant changes between pre- and post-means on several MMPI scales. The first of these was a decrease in the *F* score (p = .01); the second was an increase in the *K* score (p. = .10). Here, too, a sizable decrease occurred on the *D* score (p = .05). In contrast with the therapy group a significant decrease also occurred on the *Hy* scale (p = .05). A small gain was registered on the *Es* scale (p = .05). Again using literal interpretations, these changes may be summed up as a decrease in the psychological confusion reflected in *F*, an increase in psychological defensiveness or conversely a decrease in self-criticism as reflected in *K*, an increase in morale and optimism or conversely a decrease in depressive attitudes as reflected in *D*, a decrease in systemic complaints and the like as reflected in the *Hy* score, and a gain in ego strength as reflected in *Es*. At the same time, there were statistically insignificant decreases on all other MMPI scales with the exceptions of *Pd* and *Sc* which showed no change.

Discussion

In terms of findings bearing on Hypothesis 1, the therapy and control groups showed statistically significant differences in movement on 5 of the 8 *Q*-sort indexes when a two-tailed test and a minimum significance level of 10% is used (see Table 3). As expected, and as graphically shown in Fig. 1, the most marked difference between movements of the two groups occurred on Index 2 which measures increase in agreement between self and ideal concepts over the experimental period. As shown in Table 2, the therapy group moved from a mean z between self and ideal before therapy of .303 (see B) to a mean z of .702 after therapy (see F), while the control group moved from a pre-treatment mean z of .406 to one of .502 after the control period. Since the difference between these gains is significant with a probability of .007, it is quite clear that increase in self-acceptance as measured by Index 2 is one systematic effect of therapy that does not occur to any significant extent without therapy. The ratio of movement for the two groups on Index 2 is four to one, indicating that a quantitative difference definitely exists and, since the movement of the control group itself does not achieve significance, it suggests that a qualitative difference may exist as well.

A second significant difference in movement between the two groups appears on Index 5 which indicates that in the case of the therapy group, the change in self-perceptions over the six-week period was congruent with the relatively more stable pre-therapy ideal and probably influenced by it. In the control group, however, the change in self-perceptions over the six-week period shows no apparent relationship to the pre-treatment ideal, either positive or negative. The presence of this particular movement in the therapy group and its absence in the control group marks the difference as primarily a qualitative one.

A third significant difference in movement between the two groups is shown by Index 6. While in both groups the ever-changing self assumed a configuration more congruent with the post-ideal than was the case with the pre-treatment self, the movement of the therapy group was significantly greater. This difference is obviously a quantitative one only.

A fourth significant difference between therapy and control groups is revealed by Index 7 ($p = .02$) which indicates that over the experimental period the control group ideal moved away from the pre-treatment self while the opposite was true of the therapy group. This may be interpreted as meaning that without therapy there is a tendency ($p < .10$) to modify the ideal concept by incorporating into it values that are in direct opposition to the previously held (pre-treatment) concept of the self. This strongly suggests that the change in ideal is negatively motivated by rejection of the pre-self rather than positively motivated by carefully thinking through to a new ideal concept. The therapy group, on the other hand, went in the opposite direction—a movement not statistically significant in itself—and not only were able to accept more of the pre-therapy self, but were also able to accept new ideals that apparently resulted from something of a more constructive nature than antipathy toward the pre-self. The differences in movement on this index thus seem to be qualitative rather than quantitative.

A fifth significant difference between movements of the therapy and control groups occurred on Index 8 which may be interpreted as meaning that with therapy there is a significantly greater tendency for the changing ideal to move toward greater congruence with the newer perception of self at the termination of the treatment period. However, even though the difference in movement between groups was clear-cut ($p = .03$), both groups showed significant movement on this index: therapy group $p < .001$; control group, $p < .02$. This rather obviously indicates a quantitative rather than qualitative difference.

The five above-discussed significant differences in movement for the two groups provide an outline of the major differences in pattern of movement which rather clearly favors the therapy group. The question naturally arises as to whether these broad outlines are supported or contradicted by suggestive but less clear-cut movement.

In examining the pre-self:post-self and pre-ideal:post-ideal mean z's for the two groups (see A and H, Table 2) we find that even though the therapy group began the experimental period with less self-acceptance than the control group (B, Table 2), achieved an enormously greater gain in self-acceptance (Index 2), gained slightly more in the healthiness of the self concept (Index 1), gained slightly more in healthiness of the ideal, and generally moved farther toward a healthier personality configuration than did the control group, it manifested less change or greater stability in self and ideal concepts than did the control group. This strongly suggests that in spite of its goal of promoting change in the individual, therapy at the same time promotes stability and superimposes a pattern on the process of change which appears otherwise to be a continuous, almost random process in this population. The fact that changes in the control group are generally "favorable" even without therapy is to be expected since the over-all treatment program is aimed at producing such results. The changes in self and ideal concepts in the therapy group, then, show characteristics of clear, purposeful, integrated personality growth which do not appear in the control group.

While at first glance it may seem to the reader that convergence of the self and ideal concepts, movement of the self toward the ideal, and movement of the ideal toward the self are merely different ways of saying the same thing, such is not the case. Congruence, or self-acceptance as we sometimes refer to it, can be increased by modifying either self or ideal alone; it can be achieved by perceiving the self or the ideal in a less healthy or a more healthy fashion. It can also be achieved by simultaneous similar or dissimiliar rearrangement of the self and ideal concepts in the process of attempting to achieve congruence between the two. A difference exists in the manner in which the therapy and control groups went about achieving the increase in congruence between self and ideal concepts.

By examining the movements shown on Indexes 5 and 6 (Table 3) for the two groups, it is evident that in the process of modifying the self concept the therapy group changed in such a way that the resulting self concept was more similar to both the pre- and the post-ideal, even though the emphasis was on the latter. The control group, on the other hand, changed the self concept in such a way that the terminal self concept was more congruent with the terminal ideal but was no closer to the pre-treatment ideal than the pre-treatment self had been. The same situation holds with respect to movement of the ideal as shown on Indexes 7 and 8 (Table 3). The therapy group modified the ideal in such a way that it became more congruent with both pre- and post-self, again with the emphasis on the latter. The control group modified the ideal in a manner which indicates definite rejection of the pre-self but which resulted in greater similarity to the post-self. Therefore, it seems evident that in the control group the increase in self-acceptance was due largely to relatively superficial processes that ignore the previously held self-value system (ideal) and summarily reject previously held self-percepts and attitudes. In order to retain its gain in self-acceptance, it would appear that the control group must continue to reject its previously held self-percepts. Such a process seems similar to that described by Dymond (1955), for spontaneous remission without psychotherapy. Such a method of gaining self-acceptance seems foredoomed to failure since its success depends on continuous suppression of previously held self-percepts. In the therapy group, while much of the gain in self-acceptance is also based on new self-perceptions and new values, the change appears to be firmly integrated with and based upon the previous self-perceptions and the prior value system, i.e., the new is firmly tied in with the old, a process clearly suggesting sound personality growth.

To sum up the analysis of changes as shown by *Q*-sort data we may say that the therapy group movements present a picture of healthy, well-integrated therapeutic growth over the therapy period, a picture supported by a closer look at some of the less conspicuous relationships. The control group, on the other hand, presents a picture (see Fig. 1) of improvement to a much lesser extent on many of the individual dimensions, and furthermore appears to have gained minimally in self-acceptance only by a defensive maneuver that appears to be unstable at the outset.

As previously mentioned, only on the *Pa* scale of the MMPI did the therapy group show a significantly greater change than the control group ($p = .05$). Another significant result of therapy, then, seems to be a reduction in suspiciousness, oversensitivity, and feelings of persecution. This particular movement in the therapy group has a chance probability of less than .001, demonstrating that it is a sure, clear movement. This change is one that could be readily predicted from a knowledge of the communicative nature of group therapy. That it does not occur in the control group as well, even though all the alcoholics live and work together during their stay here, suggests that attitudes toward others are definitely reworked in the group therapeutic process but are relatively little influenced by enforced close contact with others in the hospital or by the general treatment program. Going one step further and linking this movement to the *Q*-sort movements provides us with a picture of related change familiar to any therapist. It demonstrates that perception of self and perception of others are dynamically related and tend to change in an interrelated fashion.

The picture of greater improvement in the therapy group is also supported by smaller, but still suggestive movements on the other MMPI scales as shown in Table 4.

In consideration of the evidence presented, we must conclude that the addition of Rogerian group-centered therapy to the general treatment program for alcoholics does result in both qualitative and quantitative differences in therapeutic change to a significant extent compared to a control group. Hypothesis 1 is firmly supported.

Results

Hypothesis 2

It should be kept in mind when reading the material presented in this section that the 15-session (15s) therapy group from the first phase of the project and the 30-session (30s) therapy group from the second phase of the project may differ in ways that will bias the results somewhat in one direction or another. The elapsed time between pre- and posttesting was 6 weeks in both cases. As shown in Table 1, the mean age of the 15s group was about 9 years younger, and none were over 45. The 15s group had not been drinking as long, nor had drinking been a problem for them as long, as with the 30s group. These differences appear to favor the 15s group.

Hypothesis 2: Doubling the number of therapeutic sessions from 15 to 30 will result in significantly greater therapeutic movement.

TABLE 5

Mean *z*'s for 30s and 15s Therapy Groups

Code	Sorts Correlated	30s (N=28)	*SD*	15s (N=15)	*SD*	SE_{MD}	*d*	*p*[a]
A	Pre S—Post S	.562	.225	.529	.273	.080	.033	N.S.
B	Pre S—Pre I	.303	.299	.342	.375	.108	−.039	N.S.
C	Pre S—Post I	.349	.293	.332	.252	.092	.017	N.S.
D	Pre S—C-S	.349	.263	.318	.263	.086	.031	N.S.
E	Post S—Pre I	.434	.314	.313	.306	.169	.121	N.S.
F	Post S—Post I	.702	.519	.498	.298	.149	.204	N.S.
G	Post S—C-S	.507	.265	.427	.316	.093	.080	N.S.
H	Pre I—Post I	.695	.263	.477	.332	.095	.218	<.01
I	Pre I—C-S	.654	.243	.573	.247	.078	.081	N.S.
J	Post I—C-S	.738	.234	.566	.343	.091	.172	<.06

[a] Two-tailed test of significance.

TABLE 6

Q-Sort Movement Index Means for 30s and 15s Therapy Groups

	1 G−D	2 F−B	3 H−A	4 J−I	5 E−B	6 F−C	7 C−B	8 F−E
	Therapy 30s (N=28)							
M_D	.158	.399	.133	.084	.131	.353	+.046	.268
p	<.01	<.001	<.01	<.05	<.01	<.001	N.S.	<.001
SD_D	.264	.470	.222	.198	.243	.406	.175	.334
	Therapy 15s (N=15)							
M_D	.108	.156	−.052	−.007	−.029	.166	−.010	.185
p	<.10	<.10	N.S.	N.S.	N.S.	<.01	N.S.	<.001
SD_D	.201	.286	.232	.285	.182	.204	.205	.166
	Movement Differences Between Groups							
M_D	.050	.243	.185	.091	.160	.187	+.056	.083
p	N.S.	<.08	<.01	N.S.	<.03	<.10	N.S.	N.S.
SE_{MD}	.080	.136	.074	.076	.073	.114	.061	.094

Results of Q Sort

The mean z's for the 30s and 15s therapy groups for the 10 correlations determined for each S are presented in Table 5. Index means and mean differences in movement for the two groups are presented in Table 6.

The 30s therapy group has been rather thoroughly discussed in terms of *Q*-sort and MMPI changes and there is no point in repeating it. However, a brief discussion of changes manifested by the 15s group is in order before discussion of differences in movement or pattern.

As Table 6 indicates, the 15s therapy group achieved significant movement on four of the indexes. The positive movement on Index 1 indicates that the 15s group achieved a healthier perception of the self; Index 2 reveals that it also gained greater self-acceptance by reducing the discrepancy between self and ideal during the course of therapy. A third significant change occurred on Index 6 meaning that the self moved definitely toward the post-ideal; the final significant movement occurred on Index 8 which shows that in the course of therapy the ideal was modified in the direction of the changing self. The negative movement on Index 3 indicates that for the 15s group the ideal was less stable than the self, contrary to results usually reported in the literature. Furthermore, Index 4 shows that the ideal, unlike the self, did not tend to get healthier even though it underwent considerable modification. The movement of the ideal appeared to be independent of the pre-self pattern as shown by lack of movement on Index 7, that is, movement was not based on rejection of the pre-self to any extent. The negative change on Index 5 indicates that the changing self moved slightly away from the pre-ideal, but change in the self was relatively uninfluenced by it.

The impact of therapy on the 15s group is nearly equally divided between the self and the ideal. The increased similarity of self and ideal concepts was achieved by coordinated modification of both self and ideal as shown by movement of similar amounts on Indexes 6 and 8 with greater emphasis on Index 8 which indicates that the ideal yielded more than the self. These movements at the same time resulted in a healthier self-perception without significantly affecting the relative healthiness of the ideal which nevertheless underwent considerable change in the process. In sum, the mean group movement resulted in generally favorable change.

MMPI Results

The 15s therapy group demonstrated no significant change from pre to post on any of the MMPI scales. Pre- and post-means as well as differences on the scales are presented in Table 7. Using a two-tailed test, the only significant difference in movement between the two groups occurred on the *Pa* scale ($p = .042$).

Numerous significant differences do exist between the pre-means of the two groups and between the post-means as well. As shown in Table 7, the 15s group had lower means before therapy on every MMPI variable, excepting *Es*, than the 30s group, and a higher *Es* score. This suggests that the 15s group was somewhat healthier to start with as far as scale elevation is concerned. One would, therefore, expect less change on the MMPI scales simply because the 15s group needed to move less, and, indeed, had less room in which to move since they began closer to the normal means for the scales.

Discussion

In interpreting the results bearing on Hypothesis 2, we find ourselves in somewhat of a dilemma. We would ordinarily expect

TABLE 7

MEANS AND MEAN DIFFERENCES ON MMPI SCALES FOR 30S AND 15S THERAPY GROUPS

Scale	Group	Pre-Mean	Post-Mean	M_D	p
L	30s ($N=28$)	48.607	48.107	.500	
	15s ($N=15$)	45.133	45.267	−.134	
F	30s	56.821	53.429	3.392	<.02
	15s	52.200	53.133	−.933	
K	30s	51.357	54.179	−2.822	<.10
	15s	49.933	49.800	.133	
Hs	30s	55.035	52.892	2.143	
	15s	47.200	46.533	.667	
D	30s	63.821	58.785	5.036	<.01
	15s	59.066	57.133	1.933	
Hy	30s	57.857	56.964	.893	
	15s	52.600	50.466	2.134	
Pd	30s	69.571	68.071	1.500	
	15s	64.400	64.133	.267	
Mf	30s	57.321	56.571	.750	
	15s	51.000	50.600	.400	
Pa	30s	61.250	55.250	6.000	<.001
	15s	53.266	52.066	1.200	
Pt	30s	61.607	57.000	4.607	<.02
	15s	53.400	53.133	.267	
Sc	30s	54.750	52.107	2.643	
	15s	49.000	49.866	−.866	
Ma	30s	57.678	57.750	−.072	
	15s	54.333	57.266	−2.933	
Es	30s	41.392	44.250	2.858	<.01
	15s	42.933	44.467	1.534	

30 sessions of therapy to result in more movement in the same direction shown in 15 sessions and, therefore, would use a one-tailed test of significance. However, since the possibility of qualitative differences exists we must in all honesty use a two-tailed test of significance. Inspection of Tables 6 and 7 reveals that reversals occur only where the 15s group has made no significant mean movement; for example, on Indexes 3, 4, 5, 7, and MMPI scales *L*, *F*, *K*, and *Sc*. The result is that we must use two-tailed tests of significance.

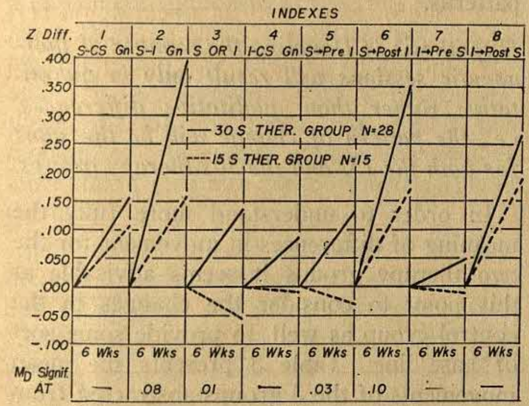


FIG. 2. Comparison of changes in 30s therapy group and 15s therapy group.

As graphically represented in Fig. 2, the most obvious difference in movement occurred on Index 2 where the probability is .08 for a two-tailed test. The 30s group gained far more in self-acceptance with a mean gain of .399 *z* while the 15s group gained .156 *z*. Self-acceptance, then, is one aspect of personality that shows continued and unabated progressive therapeutic change as the number of therapy sessions is doubled even though elapsed time is held constant. That a quantitative difference in movement exists is clear.

A large significant difference in movement also is shown on Index 3 ($p = .01$ for two-tailed test) which means that while the 15s group had a tendency to modify the ideal more than the self, the 30s group reversed this trend and rather clearly emphasized modification of the self. While this certainly suggests that a qualitative difference exists here, we must examine more of the pattern. Index 6 reveals another significant difference in movement between the two groups ($p = .10$, two-tailed test) which shows that the 30s group modified the self considerably more in the direction of the post-ideal than did the 15s group. Index 8, which indicates the extent to which the ideal was modified in the direction of the post-self, also shows a greater increase for the 30s group, though here the differences in movement were not significant.

Both therapy groups, in other words, achieved greater self-acceptance by coordinated modification of both the self and ideal concepts. However, the emphasis was different. The 30s group tended to emphasize the modification of the self while the 15s group tended to emphasize modification of the ideal to bring about greater self-acceptance.

Index 5 reveals another significant difference in movement between the two groups ($p = .03$, two-tailed test). Here, again, the 30s group shows a reversal of the direction manifested by the 15s group. The 30s group so modified the self in the course of therapy that it grew more similar to the pre-ideal as well as considerably closer to the post-ideal. The 15s group, on the other hand, modified the self away from the pre-ideal configuration but toward the post-ideal pattern. At the same time, the 30s group reversed the 15s pattern on Index 7 which indicates that while the 15s group tended to modify the ideal away from the pre-self though toward the post-self, the 30s group tended to modify the ideal to include more of the pre-self as well as more of the post-self.

Indexes 1 and 4 show that the average gain in healthiness of both self and ideal when compared to the criterion-sort (C-S) was slightly but not significantly larger for the 30s group. As might be expected from the foregoing, the 15s group showed no improvement in the healthiness of the ideal concept.

The fact that the 15s group had a more normal group profile to begin with makes it difficult to find much evidence relevant to

Hypothesis 2 in the MMPI comparisons. However, there was considerable room for improvement on the *D* scale, for example, yet the 15s group showed less than half as much reduction as the 30s group. In general, the 30s group showed greater mean movement than the 15s group where such comparisons are justified.

It should be pointed out that, although the 15s group had a more nearly normal mean profile before therapy than the 30s group, they had no such advantage on the *Q*-sort measures. On the latter measures (see Table 5) the two groups showed reasonable similarity. Therefore, the much smaller movements of the 15s group on *Q*-sort indexes can scarcely be attributed to lack of room for improvement.

A final glance at Table 5 brings to light a most interesting finding. In spite of the much greater therapeutic movement shown by the 30s group, it showed less change in either the self or ideal concepts than did the 15s group (see A and H). This phenomenon is similar to that discussed in the comparison of the 30s group and the six-week control group.

In consideration of the evidence, we must conclude that Hypothesis 2 is supported: doubling the number of therapeutic sessions from 15 to 30 results in significant gains in therapeutic movement to a significant degree, though not on all variables measured, even though elapsed time for the course of therapy was essentially similar (5 weeks for 15s group; 6 weeks for 30s group), and time between pre- and posttests was identical (6 weeks).

RESULTS

Hypothesis 3

At this point it is rather obvious that while doubling the number of therapy sessions results in a significant increase in therapeutic movement, something more than mere quantitative change seems necessary to account for the reversals of direction and the shifting emphases. In other words, on a quantitative basis alone we have considerable evidence that Hypothesis 3 is untenable. It is still deserving of separate consideration because of light it may shed on the nature and process of therapeutic change. The results bearing on this hypothesis are the same statistics used in testing Hypothesis 1 and 2 and, therefore, appropriate tables will be referred to rather than duplicated. The conceptual framework within which these data are evaluated, however, will differ from that used to test Hypothesis 2. Only *Q*-sort data will be used since the other measures do not seem sensitive enough to short-term change to yield meaningful patterns.

Hypothesis 3: Doubling the number of therapeutic sessions will result only in quantitative rather than qualitative differences, i.e., the pattern of change will be the same for both the 15s and the 30s therapy groups.

In order to understand more fully the meaning of differences in movement for the two therapy groups it seems advisable at this point to consider the changes in the control group as well, to provide some sort of base line. Table 8 presents the mean movements of the 3 groups abstracted from Tables 3 and 6 for convenience in following the discussion. Figure 3 represents this in graphic form.

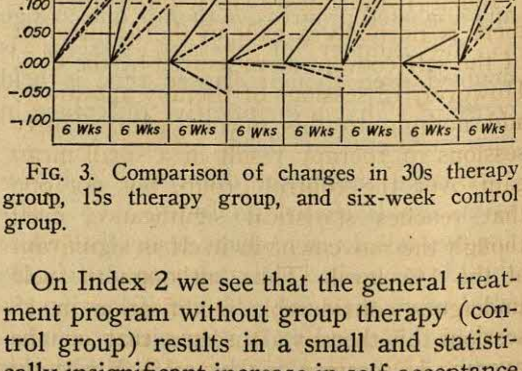


FIG. 3. Comparison of changes in 30s therapy group, 15s therapy group, and six-week control group.

On Index 2 we see that the general treatment program without group therapy (control group) results in a small and statistically insignificant increase in self-acceptance of .096; 15 sessions of therapy resulted in

TABLE 8

MEAN MOVEMENT ON *Q*-SORT INDEXES OVER THERAPY OR CONTROL PERIOD FOR 30S AND 15S THERAPY GROUPS AND FOR CONTROL GROUP

Group	*Q*-Sort Index							
	1	2	3	4	5	6	7	8
30s	.158	.399	.133	.084	.131	.353	+.046	.268
15s	.108	.156	−.052	−.007	−.029	.166	−.010	.185
Control	.114	.096	.055	.042	−.007	.192	−.096	.103

a movement significant at the 10% level and more than 1½ times the movement of the control group; 30 sessions of therapy resulted in a movement with a probability value of .001 and more than double the amount of movement of the 15s group; more than four times the movement without therapy.

On Index 3 we see that without therapy there is a slight tendency for the self to be less stable than the ideal. However, with 15 sessions of therapy there is a tendency for the ideal to be less stable than the self. With 30 sessions there is a clear concentration of change on the self.

Index 4 shows that without therapy (control group) the ideal actually tends to increase in adjustment as measured by the C-S. With 15 sessions of therapy, however, the ideal has been modified to the point where it is becoming less adjustive. With the larger number of therapy sessions a clear gain in the adjustment of the ideal is again achieved.

Index I indicates that without therapy there is normally a gain in the adjustment of the self concept as measured by the C-S. However, 15 sessions of therapy appears to add nothing to this normal gain. Thirty sessions of therapy result in a small mean gain over the control group, but not one that reaches statistical significance, even though the movement in itself is significant at the 1% level. Thus, although the self undergoes considerable modification in 15 sessions of therapy, it seems that on the average 15 sessions of therapy does nothing to improve the adjustment of the self concept and longer-term therapy is needed before improvement in healthiness of the self concept that may be properly ascribed to therapy actually begins to occur.

Index 5 shows that without therapy change in the self has little tendency to move either toward or away from congruence with the pre-ideal. With the shorter number of therapeutic sessions the self tends to drift away from the pre-ideal but moves toward it with 30 sessions.

Indexes 6 and 8 indicate that without therapy most of the gain in self-acceptance is brought about by modifying the self in the direction of the post-ideal and less by modifying the ideal in the direction of the post-self. In the 15s group the movement is more nearly balanced but the emphasis is on modifying the ideal in the direction of the self. With the 30s group the emphasis once more is on modifying the self toward the emerging ideal.

Finally, Index 7 indicates that without therapy, there is a significant tendency to modify the ideal in a manner suggestive of rejection of the pre-self. The 15s group has succeeded in halting this trend, while the 30s group has managed to reverse it.

Discussion

Results presented so far indicate that the 15s and the 30s therapy groups show the following similarities in pattern:

1. The self concept has grown significantly healthier.

2. Self-acceptance has significantly increased.

3. The self has moved significantly toward the modified ideal.

4. The ideal has moved significantly toward the modified self.

But the following dissimilarities appear:

1. The 15s group ideal was less stable than the self; the opposite is true of the 30s group.

2. The 15s group ideal did not gain in health while that of the 30s group did.

3. The 15s group self tended to move away from rather than toward the pre-ideal; the opposite was true of the 30s group.

4. The 15s group tended to modify the ideal in a manner suggesting the possibility of a rejection of the pre-self while the 30s group modified the ideal in a manner suggesting greater acceptance of the pre-self.

5. In the process of increasing self-acceptance, the 15s group modified the ideal more in the direction of the changing self than it modified the self in the direction of the ideal, an emphasis reversed by the 30s group.

It is quite clear at this point that if we consider 15 sessions and 30 sessions of therapy as two examples of therapeutic process carried through to completion in both cases, differences in pattern for the two groups do exist and Hypothesis 3 must be rejected. However, when we take the movements of the control group into consideration, it is apparent that these differences cannot be due simply to the possibility that the 15s group had not had enough therapy to influence these measures; the differences seem to result from therapy, not from lack of it. Why these particular differences arise must somehow be related to the dynamics of therapeutic change, but data presented so far does not provide a sufficiently clear picture for understanding them. At this point we may, on the evidence presented, firmly reject Hypothesis 3. We will return to it later after we have explored Hypothesis 4 in the hope that we will then be in a better position to understand the process of therapy.

RESULTS

Hypothesis 4

Data bearing on Hypothesis 4 are again taken from *Q* sorts performed by the 30s therapy group immediately preceding the initiation of therapy, at the end of the 10th session, at the end of the 20th session, and at the end of the 30th session. For convenience, these are hereafter referred to as the 0, 10, 20, and 30 sorts, *0* corresponds to *pre* in previous sections, *30* corresponds to *post.*

Hypothesis 4: As the number of therapy sessions increase, the rate of therapeutic progress will increase.

Results of Q Sort

To evaluate this hypothesis we will again use the indexes to trace therapeutic progress. Since we were concerned, initially, with progressive movement from the pre-therapy zero point, indexes were calculated for the 0s–10s, the 0s–20s, and the 0s–30s period. These values are given in Table 9, and indicated graphically in Fig. 4.

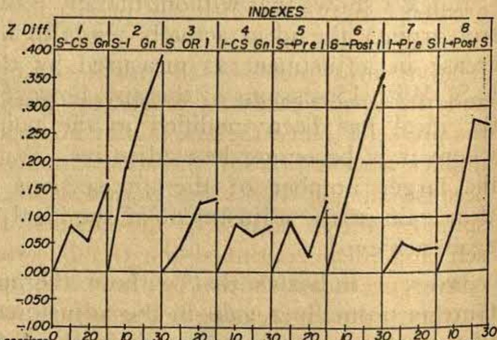


FIG. 4. Movement of 30s therapy group calculated from pre-therapy configurations for the 0s-10s, 0s-20s, and 0s-30s.

Examination of Figure 4 readily reveals that none of the individual indexes support Hypothesis 4 and furthermore that Indexes 1, 4, 5, 7, and 8 show regressive movement. Only Indexes 2 and 6 show a steady upward climb; Index 3 rather suddenly seems to level off. Indexes 1 and 5 are particularly

TABLE 9

SIGNIFICANCE OF MOVEMENT OF 30s THERAPY GROUP AT DIFFERENT PERIODS

	1	2	3	4	5	6	7	8
$0\text{-}10_M$	.082	.182	.046	.087	.091	.130	.052	.091
SD	.170	.352	.239	.187	.169	.268	.179	.221
p	<.02	<.02	N.S.	<.05	<.01	<.02	N.S.	<.05
$0\text{-}20_M$	.055	.307	.122	.067	.030	.267	.040	.277
SD	.276	.529	.269	.178	.279	.468	.151	.363
p	N.S.	<.01	<.05	<.10	N.S.	<.01	N.S.	<.10
$0\text{-}30_M$	.158	.399	.133	.084	.131	.353	.046	.268
SD	.264	.470	.222	.198	.243	.406	.175	.334
p	<.01	<.10	<.01	<.05	<.01	<.10	N.S.	<.10

interesting because, as shown in Table 9, both show significant movement during the first two weeks of therapy but when calculations are based on the 0s–20s period, the mean gains are no longer significant. But by the end of the 0s–30s period the over-all gains are again significant. In other words, during the middle two weeks of the six-weeks therapy period there occurs a loss of relative adjustment of the self concept and this is accompanied by a temporary reversal of the trend of the self concept toward greater congruence with the pre-ideal. At the same time there is a slight loss in the relative adjustment of the ideal concept; and the ideal ceases its movement toward the pre-self. Nevertheless, this by no means implies a cessation of therapeutic activity because during the same period, and in spite of these regressive movements, the self and ideal continued the trend toward convergence, as shown by Index 2, and further, the increasing congruence was brought about by approximately equal movements of self toward post-ideal and ideal toward post-self as shown in Table 9.

In the final two weeks of the therapy period convergence of self and ideal continues, but during this period the relative adjustment of the self again shows a significant gain which is followed by the ideal, although the latter gain is not significant. During this period the self also resumes its movement toward the pre-ideal and continues its movement toward the post-ideal. The ideal, on the other hand, seems to have lost its momentum toward either the pre-self or the post-self. Hypothesis 4 must be rejected; increasing rates of change are nowhere in evidence on these variables.

A tentative interpretation of the various regressive movements in the context of progressive convergence of self and ideal concepts is that the means by which convergence is initially brought about results in an unstable configuration, necessitating certain changes before better integrated movement can occur.

We may also look at therapeutic change from a slightly different point of view and examine the nature of the changing relationships within each two-week segment, basing our calculations on the three sets of initiation and termination points, 0s–10s, 10s–20s, and 20s–30s, independently of each other. Movements on Indexes 1, 2, and 4 are necessarily additive for each index because of the manner in which they are computed, but movements on the other indexes are not and hence may throw further light on the process.

Figure 5 and Table 10 present the results of computing the indexes for each of the two-week periods of therapy independently. Values for the 0s–10s period are identical to those already given, of course, and are repeated only for the sake of comparison and completeness. The most notable feature in the 0s–10s period is that all movements

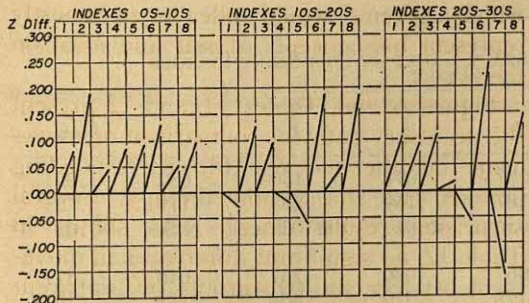


FIG. 5. Movement of 30s therapy group calculated independently for each two-week period.

are positive and, with the exception of Indexes 3 and 7, are significant.

During the second two-week segment taken independently, three of the indexes show regressive movement compared to four with regressive movement when progress was computed from the pre-treatment configuration (see Fig. 4 and Table 9). The disagreement occurs on Index 7 and may be interpreted as meaning that while the modification of the ideal in the direction of the pre-treatment self (at 0s) was abandoned after the first two weeks of therapy, the process continued during the second two-week period but with different content: during the second period the ideal was evolving toward greater congruence with the self concept that had developed at the end of the first two weeks of therapy (at 10s). At the same time, i.e., during the 10s–20s period, the changing self tended to move away from the ideal configuration that had developed over the first 10 sessions of therapy, as shown on Index 5, Fig. 5, and lost much of its increase in congruence with the pre-treatment ideal (at 0s) as shown on Index 5, Fig. 4. Nevertheless, congruence between concurrent self and ideal concepts continued to increase in the second two-week segment (see Index 2, Fig. 5). This was achieved by approximately equal and concomitant modifications of the self and ideal concepts toward each other (Indexes 6 and 8, Fig. 5). Maintaining progress toward greater congruence of self and ideal concepts during the middle two-week period was possible only by a sacrifice in the relative adjustment of both the self and ideal concepts (Fig. 5, Indexes 1 and 4).

During the final two-week period, the self concept recovered its loss in relative adjustment and progressed beyond the earlier peak. The ideal did not quite fully recover its loss by the end of the sixth week, although the difference was infinitesimal. A possible reason for failure of the ideal to rebound as quickly as the self, as far as

TABLE 10

MOVEMENT ON INDEXES GIVING CONSIDERATION TO THE THREE 2-WEEK PERIODS INDEPENDENTLY (FOR 30S GROUP)

	1	2	3	4	5	6	7	8
0-10$_M$	.082	.182	.046	.087	.091	.130	.052	.091
SD	.170	.352	.239	.187	.169	.268	.179	.221
p	<.02	<.02	N.S.	<.05	<.01	<.02	N.S.	<.05
10-20$_M$	−.027	.125	.096	−.020	−.060	.174	.049	.185
SD	.263	.525	.232	.154	.223	.449	.193	.390
p	N.S.	N.S.	<.05	N.S.	N.S.	<.06	N.S.	<.05
20-30$_M$	.103	.092	.106	.017	−.057	.250	−.158	.149
SD	.279	.418	.275	.141	.477	.363	.386	.344
p	<.10	N.S.	<.06	N.S.	N.S.	<.01	<.05	<.05
0-30$_M$	.158	.399	.133	.084	.131	.353	.046	.268
SD	.264	.470	.222	.198	.243	.406	.175	.334
p	<.01	<.10	<.01	<.05	<.01	<.10	N.S.	<.10

relative adjustment value is concerned, may be that such a movement would interfere with progressive convergence of the self and ideal concepts.

In terms of the over-all picture, 0s–30s, the ideal made no further progress toward the self in the final two weeks of therapy (see Index 8, Fig. 4), but in terms of the final segment taken independently, the ideal held at 20s did undergo modification in the direction of the post-treatment self (at 30s) as shown by Index 8, Fig. 5. At the same time, during the final two-week period, the ideal was modified in a direction away from the less adjustive self concept that was held at 20s (Index 7, Fig. 5). Further, the 30s self had moved away from the less adjustive ideal held at 20s (Index 5, Fig. 5) but toward the ideal held at the termination of therapy (30s) as shown on Indexes 6 and 8, Fig. 5.

These data raise at least two questions. The first question is: What is responsible for the regressive movement? The second is: What is responsible for the resumption of progressive therapeutic movement? Since regression appears to occur during the 10s–20s period, further analysis of movement during the first two weeks of therapy should help provide some understanding of the phenomenon. In order to do this we must make use of our second control group. This group was selected by the same procedure as were the others and the two-week measurement period corresponds to the first two weeks of therapy (0s–10s) for the 30s therapy group. As mentioned earlier, they were selected after group therapy had been terminated for the summer to avoid the influence of the negative reactions and resentments that occur when *S*s not selected for therapy are asked to retake the tests which they may feel were responsible for their being left out of a therapy group. It is assumed that such feelings do not materially influence the results of the final tests for the six-week control group because of the additional time involved. The bias we were most concerned about here was primarily that of the effect of cooperativeness or uncooperativeness on the accuracy with which a given *S* would endeavor to honestly represent his concepts of self and ideal on the *Q* sort.

Figure 6 and Tables 11 and 12 present the necessary data for comparison of movement on the indexes for corresponding periods with and without therapy. Several surprising results emerge. As shown in Table 12, no significant differences in movement between the two groups are apparent on any of the indexes. Table 11 shows only two significant differences. The therapy group had a significantly more stable self concept than did the control group (see A). At the end of the two-week period, the therapy group held a significantly more adjustive ideal concept (see J).

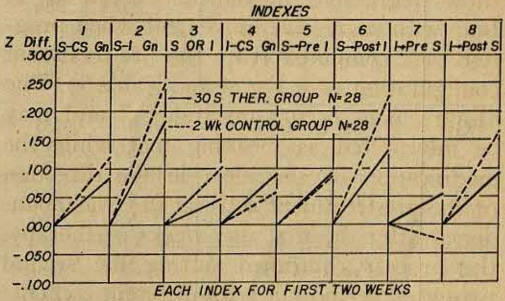


FIG. 6. Comparison of movement for corresponding periods with and without therapy.

Inspection of Fig. 6 reveals that only on Index 7 does there appear to be any indications of a qualitative difference in the movements of the two groups, and this difference is significant at about the 20% level. The fact that the ideal of the control group tended to be modified in a direction implying rejection of the pre-treatment self while the therapy group modified the ideal in a manner that resulted in achievement of greater congruence with the pre-treatment self suggests that here is one of the more subtle effects of therapy. That this difference in the process of modification of the ideal is a genuine one, rather than artifact, is indicated by the previous comparison of the six-week control group and the 30s therapy group, as shown back in Fig. 1, suggesting that these early differences in direction of movement continue throughout the entire six-week period. At the end of

TABLE 11

COMPARISON OF MOVEMENT FOR CORRESPONDING PERIODS WITH AND WITHOUT THERAPY

Code	Sorts Correlated	0-10 30s Therapy Group (N=28)	0-10 2-Week Control Group (N=28)	M_D	p
A	Pre S—Post S	*M* .655 *SD* .175	*M* .525 *SD* .191	.130 SE .050	<.01
B	Pre S—Pre I	*M* .303 *SD* .299	*M* .358 *SD* .294	−.055 SE .081	N.S.
C	Pre S—Post I	*M* .355 *SD* .276	*M* .383 *SD* .241	−.028 SE .070	N.S.
D	Pre S—C-S	*M* .349 *SD* .263	*M* .402 *SD* .222	−.053 SE .066	N.S.
E	Post S—Pre I	*M* .394 *SD* .286	*M* .444 *SD* .287	−.050 SE .078	N.S.
F	Post S—Post I	*M* .485 *SD* .396	*M* .608 *SD* .342	−.123 SE .101	N.S.
G	Post S—C-S	*M* .431 *SD* .313	*M* .520 *SD* .209	−.089 SE .072	N.S.
H	Pre I—Post I	*M* .701 *SD* .270	*M* .622 *SD* .311	.079 SE .079	N.S.
I	Pre I—C-S	*M* .654 *SD* .243	*M* .577 *SD* .279	.078 SE .071	N.S.
J	Post I—C-S	*M* .741 *SD* .217	*M* .633 *SD* .207	.108 SE .058	<.06

TABLE 12

COMPARISON OF INDEXES FOR CORRESPONDING PERIODS WITH AND WITHOUT THERAPY

	1	2	3	4	5	6	7	8
	30s Therapy Group							
0-10$_M$	.082	.182	.046	.087	.091	.130	.052	.091
SD	.170	.352	.239	.187	.169	.268	.179	.221
p	<.02	<.02	N.S.	<.05	<.01	<.02	N.S.	<.05
	2-Week Control Group							
0-10$_M$	.118	.250	.097	.056	.086	.225	−.025	.164
SD	.195	.313	.266	.218	.186	.257	.192	.248
p	<.01	<.01	<.10	N.S.	<.05	<.10	N.S.	<.01
M_D	−.036	−.068	−.051	.031	.005	−.095	.077	−.073
p	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.

the six-week period a difference significant at the 2% level exists, when a two-tailed test is used.

Figure 6 also hints of other differences in change, even though the differences are of no statistical significance, for the initial two-week period. In general, Fig. 6 reveals that the control group made greater gains on five of the indexes than did the therapy group. Ignoring the possibility of sampling error for the moment, we may say that one apparent effect of group therapy is that of retarding movement, at least in the initial period of therapy.

That regressive movement of considerable proportions must also occur in the control groups is suggested by the fact that at the end of the six-week period the therapy group showed significantly greater gains on five of the indexes (see Fig. 1). Four of these (Indexes 5, 6, 7, and 8) deal with the manner in which the *S*s went about modifying the self and ideal concepts in order to achieve greater congruence, and the fifth one (Index 2) measures the increase in congruence attained. Since the therapy group evidently gained self-acceptance at a slower pace and by a somewhat different route, and because the therapy group managed to maintain progress toward greater self-acceptance, while the control groups did not—if we can regard the two control groups as equivalent samples—then it appears that the stability of the increase in self-acceptance is largely dependent upon the means used to achieve it. The most obvious clue to the nature of this difference is provided by Index 7, which shows that both the two-week and the six-week control groups (see Fig. 6 and 1) sought to achieve greater self-acceptance by modifying the ideal in a manner implying a rejection of the pre-treatment self. The six-week therapy group, on the other hand, at no time resorted to this particular method, as far as the group mean is concerned, with respect to the pre-treatment self, although as shown on Fig. 5 it did occur with respect to the self concept held at 20s.

These comparisons still do not provide an answer to either of our two questions posed earlier, viz., what produces the regressive movement and what is responsible for the resumption of progressive movement. From the foregoing comparisons, it seems reasonable to assume that the control group also underwent regressive movement, but did not recover, since the two-week control group revealed, in general, greater movement than did the six-week control group.

In an attempt to provide a tentative answer to the "Why?" of regressive movement in the therapy group we must rely on clinical experience. The increase in the relative adjustment of the self and ideal concepts early in therapy with a concomitant gain in self-acceptance appears to represent a phenomenon usually called a "flight into health." That it represents such a defensive reaction is strongly suggested by the fact that both self and ideal have begun to regress, in terms of relative adjustment, at some point after the tenth session. This regression is probably brought on by a growing awareness of the artificiality of the improvement in conjunction with continued therapeutic pressure in the group. That this early "peaking" represents "flight into health" rather than truly integrated therapeutic change is supported by a wealth of clinical observation. It was at this point, allowing for individual differences of course, that *S*s tried hardest to leave therapy. Apparently individuals who have fled into "health" as a defense against self-examination and criticism have not completely fooled themselves (or us either) because they seem to exhibit a sense of impending doom, high anxiety, mildly hypomanic behavior, extreme rationalization, and flat-footed lying in order to convince others, and themselves perhaps, that all is well and that they are ready for discharge, have "seen the light," etc. Some alcoholics go to extraordinary lengths to try to terminate commitments, some even escape. Often the less imaginative and less talented will explain their efforts by saying that they feel that they must be released before they loose what they have already gained. On the average, this occurs at about the middle of the 60-day program which corresponds to about the tenth to fifteenth therapy session. This phenomenon seems to occur with or without therapy. Our data in Fig. 4 and Table 9 indicate that there is good reason for the feeling that what has been gained may be soon lost. Just where in the 30-session span the low points for self and ideal are reached we cannot tell from our data other than that in the therapy group it falls somewhere between the tenth and twentieth sessions. Here the 15s group can help us to some extent when we recall that mean gain on Index 1 (self:C-S) was a little less than that of the control group; there was a slight loss in relative adjustment of the ideal compared to a gain for the control group, and ideal had changed more than self while the reverse was true for the control group. Compared to the 30s group, the 15s group had gained less than half as much in self-acceptance (see Fig. 3).

If we are permitted the liberty of superimposing the pattern of movement of the 15s group on the pattern of the 30s group, it would indicate that the self:C-S correlation must reach a peak considerably higher shortly after the tenth session and then begin to fall. The ideal:C-S, to account for our 15s results, must peak about the same time as does the self:C-S but apparently regresses much more rapidly, to a point even below its pre-therapy level. In terms of our 30s group pattern, this interpretation would mean that lowered correlations at the twentieth session are points on a recovery phase rather than true low points in the process.

A thing not easily explained is why, in spite of the ups and downs of the relative adjustment of the self and ideal concepts, self-acceptance continues to grow. With roughly parallel movements of self and ideal when compared to the criterion, one would expect that self-acceptance would remain constant, but this is not the case. A possible explanation is that the need for self-acceptance is the driving force, especially after the trauma of being hospitalized for alcoholism and possibly an additional jolt at being put in a therapy group, or we may simply label it a desperate attempt to reduce a considerable load of anxiety in a situation where alcohol is not available to perform the service. If self-acceptance is the driving force or "felt-need," then it becomes understandable why it increases so rapidly and simultaneously with increased healthiness of both self and ideal. Given the need for self-acceptance and given many new ideas in ways to look at one's self and at the same time given an estimate of the social acceptability and possible value to the individual in behaving and perceiving and striving in certain ways by the orientation lectures and the many AA functions, as well as the experience of being in, for the alcoholic, an unusually friendly and accepting environment, it is rather easy to understand how the individual may begin to believe that many of the things he hears about the "average alcoholic" are indeed true about him and furthermore that these things are socially acceptable and, therefore, he wants them to be true about himself. If this is indeed the case, the large increase in self-acceptance follows automatically as far as our methods of measurement are concerned. However, as previously mentioned, anxiety as inferred from clinical observation is higher at this point, not lower as it should logically be if self-acceptance has really grown to the extent indicated by our statistics. This can mean just one thing, that the initial large increase in integration is just as superficial as is the initial increase in relative adjustment of the self and ideal concepts. While our data do not show it, it is quite possible that self-acceptance also undergoes a change for the worse at some point between the tenth and twentieth sessions. However, because loss of self-acceptance tends to increase anxiety, even though the individual may be losing something very superficial from a clinical point of view, he is apt to exert considerable effort to prevent it. A possible explanation of how this might come about is that the individual begins to realize his self concept is unrealistic, an awareness aggravated by group therapy; in order to prevent increased anxiety through facing the realization that he is falling far short of his newly acquired ideals, he begins to jettison them as they become sources of anxiety with awareness that in reality he falls far short of them, instead of being a source of self-satisfaction as they were when they were so readily accepted shortly before. By rejecting ideals as rapidly as they begin to hurt, the relative adjustment of the ideal is decreased as rapidly as is the relative adjustment of the self as realism, nurtured by therapy, begins to creep into his self concept. As might be expected, the ideal drops much more rapidly, and much farther if we use results of the 15s group as an indication, possibly a matter of overcompensation for what in some *S*s almost amounts to panic. Clinically, depression of some degree is usually manifest along with anxiety.

For those in group therapy, this is typically the point at which serious therapeutic work begins. The whole artificial structure seems to come down with a resounding crash, and with the help of the group and the therapist the *S* begins to pick up the pieces and put them together in a process of sound therapeutic growth. During this period, according to the data in Table 10 and Fig. 5, the self concept undergoes the greatest gain in health as the individual examines himself more honestly, ideals are carefully thought through and the level of adjustment of the ideal rises beyond its initial level, but not to its earlier peak, and self-acceptance is placed on a sounder footing in the process. It is our belief that it is during this final phase that genuine therapeutic integration and growth take place.

If this picture is correct, then we have a tentative answer to the second question posed regarding the "why" of the resumption of progress. Having experienced the rewards of increased self-acceptance, even though in an unrealistic manner during the flight into health, the therapy client is quite likely unwilling to give up the new-found feeling of self-satisfaction. His motivation to maintain it is probably quite high. But the pressure of group therapy prevents him from successfully maintaining self-acceptance on an unrealistic basis unless he relinquishes the benefits and satisfactions of continued group acceptance and group participation. While some so chose, most seem willing, with the support of the group, to examine themselves more penetratingly, to face realities that they could only deny earlier in therapy. Clinically, this is the period when interpersonal defenses are at low ebb, and the group begins to take on aspects of unity and solidarity. Without therapy, the alcoholic is apparently not forced to make the choice between (*a*) further painful examination to put self-acceptance on a more realistic level or (*b*) loss of group acceptance. He apparently undergoes an even greater regression but manages to salvage only a small amount of his new-found self-acceptance. Even this is not on a sound basis since, as we have seen, the resultant ideal implies rejection of the pre-treatment self concept. The difference in recovery between the therapy and control groups thus appears to be due to at least three factors.

1. The experience and help in self-examination afforded by participation in group therapy.

2. The pressure toward reality-oriented thinking in the group setting.

3. The continuing and increasing reward of group acceptance when the individual struggles to put self-acceptance on a sound basis, or perhaps the threat of rejection if he fails to do so.

At this point we may briefly reconsider Hypothesis 3, regarding pattern differences in the 15s and 30s therapy groups. In view of what has been said regarding the mean pattern of therapeutic progress, we can now say with some conviction that the differences result from the termination of therapy in the 15s group somewhere in the early stages of the reintegrative phase. From this it seems to follow that therapeutic progress depends more upon the number of therapy sessions than upon the elapsed time between the beginning and end of therapy. Where patients are hospitalized as were our *S*s, intensive psychotherapy, five days a week in this case, is definitely advantageous in making maximal use of every day of in-hospital time.

STABILITY OF SELF AND IDEAL CONCEPTS

If what we have said about the pattern of the therapeutic process is true, we would expect stability of the self and ideal concepts to be lowest over the initial period when the individual is "fleeing into health," and highest in the final phase when change is based on laborious thinking-through. Since we have apparently straddled the disintegrative low spot with our tenth and twentieth session measurements, measurements taken at the first peak and at the bottom following collapse of the artificial structure would probably show a lower stability than any of those in Table 13.

TABLE 13

STABILITY OF SELF AND IDEAL CONCEPTS AT 2-WEEK INTERVALS IN THERAPY FOR 30S GROUP

Concept	First 2 Wks. 0-10		Second 2 Wks. 10-20		Third 2 Wks. 20-30		Overall 6 Wks. (0-30)	
	z	M_D	z	M_D	z	M_D	z	M_D
Self:Self	.655	.030	.685	.054	.739		.562	
Ideal:Ideal	.701	.080	.781	.064	.845		.695	

The data in Table 13 support our expectations and add another bit of information regarding the process of therapy. This extra bit derives from inspection of the mean differences. While for both self and ideal the trend is toward increased stability, the *rates* of increase are in opposite directions. Stability of the self seems to increase at an increasing rate, but stability of the ideal increases at a diminishing rate. Stability of the self concept has been shown by Brownfain (1952) and other investigators to be one measure of psychological health and we would go even further and tentatively apply the same interpretation to stability of the ideal: we find that changes in stability follow the changes in the adjustment of the self and ideal concepts measured by the criterion sort. Thus, in therapy, there is a positive relationship between increasing the relative adjustment of the configuration and increasing the stability of that configuration. The greatest increase in the stability of the ideal occurs early in therapy when it also shows the greatest gain in relative adjustment; the greatest increase in stability of the self occurs in the reintegrative phase of therapy when it also shows the greatest gain in relative adjustment (see Table 13). That these changes represent progressive therapeutic movement rather than "random" variation is supported by the previous discussion and data.

If this interpretation is correct, what can we predict about the six-week stability figures for the 15s therapy group? We have reasoned that the 15s group was terminated in the early stages of the reintegrative phase. Since the ideal at that point was actually slightly less healthy than it was originally, due, we said, in order to maintain self-acceptance of the deteriorating self concept, we have captured the ideal at a point where it still shows the effect of having been forcibly distorted and therefore should show a low correlation with the ideal measured at nearly any other point. That is, we can predict that for the 15s group the stability figure between the pre- and post-ideal will be inordinately low, lower than the stability of the self over the same period because deterioration of the healthiness of the self concept was, we have assumed, the result of painful awareness rather than deliberate downward distortion as was the case with the ideal; furthermore, there were indications that downward distortion of the ideal was overzealous and that it actually underwent greater temporary modification than did the self. Because reintegration was under way at the termination of therapy, but had by no means been completed, we can predict that pre-post self stability will be lower than for the 30s group.

What can we predict about stability of the self and ideal concepts of the control group over a comparable period of time? Knowing that therapy tends to increase the stability of the self and ideal concepts as well as leading to greater acceptance of the pre-self as well as the post-self, while without therapy there is a tendency to reject the pre-self in the modification of the ideal (these points were discussed under Hypothesis 1), we can predict that stability will be generally lower for the control group. The data in Table 14 confirm these predictions.

TABLE 14

SIX-WEEK STABILITY VALUES FOR CONTROL GROUP, 15s AND 30s THERAPY GROUPS

Concept	Control ($N=28$)	15s ($N=15$)	30s ($N=28$)
Self:Self z	.494	.529	−.562
Ideal:Ideal z	.549	.477	.695

It should be noted that the stability of the ideal for the 15s group is actually lower than that of the control group which is, of course, understandable in light of what has been said about forcible regression of the ideal in the 15s group.

We have one final facet to explore before stating our conclusions: a comparison and interpretation of differences that may exist between movement and stability values of the 30s therapy group over the 0–10 two-week period and the two-week control group over a period of the same length but without therapy. So far we have presented no evidence, other than clinical judgments, that tells us whether or not the phenomenon referred to as "flight into health" actually does occur without therapy under the total impact of the fact of hospitalization and the stimulus of the general treatment program. The 0–10 two-week period for the 30s group corresponds to the same two-week period relative to initiation of the 60-day hospitalization period, i.e., both periods began at an average of 10 days after admission.

TABLE 15

STABILITY OF SELF AND IDEAL CONCEPTS OF 30s THERAPY GROUP FOR 0-10 TWO-WEEK PERIOD AND CORRESPONDING PERIOD FOR TWO-WEEK CONTROL

Concept	30s Group ($N=28$)	Two-Week Control Group ($N=28$)	d	p
Self:Self z	.655	.525	.130	$<.01$
Ideal:Ideal z	.701	.622	.079	N.S.

The data in Table 15 clearly indicate that group therapy actually retards variability or instability of self and ideal concepts, from which we may infer that it also retards the flight into health. This being the case, from what we have said regarding the initial spurious gain in self-acceptance during the burgeoning of this artificial structure, we should expect that the self:ideal correlation at the end of the two-week period for the two-week control group should be considerably inflated compared to the increase shown by the 30s group.

As Table 16 reveals, this is exactly what seems to be happening. Table 17 supports the argument in revealing a comparatively large increase in the healthiness of the self

TABLE 16

SELF:IDEAL CORRELATIONS BEFORE AND AFTER CORRESPONDING INITIAL TWO-WEEK PERIOD FOR 30s THERAPY GROUP AND TWO-WEEK CONTROL GROUP

Group	Self:Ideal			
	z at 0	z at End of 2nd Wk.	d	p
30s group ($N=28$)	.303	.485	.182	$<.02$
Two-week control group ($N=28$)	.358	.608	.250	$<.001$

TABLE 17

SELF-CRITERION AND IDEAL-CRITERION CORRELATIONS BEFORE AND AFTER CORRESPONDING INITIAL TWO-WEEK PERIOD FOR 30s THERAPY GROUP AND TWO-WEEK CONTROL GROUP

Group	Concept	0	End of 2nd Wk.	d	p
30s Group ($N=28$)	Self:Criterion	.349	.431	.082	$<.02$
	Ideal:Criterion	.654	.741	.087	$<.05$
Two-week control group ($N=28$)	Self:Criterion	.402	.520	.118	$<.01$
	Ideal:Criterion	.577	.633	.056	N.S.

concept. The ideal, however, shows less increase than that of the therapy group. However, since the ideal of the two-week control group manifested lower stability over this period than that of the therapy group as shown in Table 15, much of the change in the ideal undoubtedly reflects the modification of the ideal to fit the self as part of the process of flight into health.

The phenomenon of "flight into health," then, is by no means confined to those in therapy groups, but is rather a general characteristic of the alcoholic population hospitalized for treatment and undergoing the general treatment program at Willmar State Hospital. Group therapy apparently has a sobering effect on the feverish drive for self-acceptance in the initial phase of treatment and, after the inevitable collapse, provides the means, the emotional support, and perhaps much of the motivating force necessary to pick up the pieces and laboriously and painfully build a healthier, sounder phenomenological world. As shown by results presented under Hypothesis 1, without therapy this reintegrative process does not occur and the net gain as a result of hospitalization in terms of therapeutic growth as we define it and measure it is minimal.

It should be clear from our discussion and generalized description of the flight-into-health phenomenon, that with or without therapy, the worst possible time for an alcoholic to be discharged is the period after this process has begun but before it has collapsed from its own weight or therapeutic pressure, as the case may be, since a period of intense anxiety and depression is imminent and if it occurs shortly after discharge the chances would seem to be extremely good that the alcoholic will resort to that old standby, alcohol, to alleviate the pain. Discharge while the individual is at or near the bottom of the trough might be even more disastrous even sooner.

CONCLUSIONS

1. The addition of Rogerian group-centered therapy to the general treatment program for alcoholics does result in both qualitative and quantitative differences in therapeutic change to a significant extent. These differences are: (*a*) A marked increase in self-acceptance. (*b*) The emphasis of change was clearly centered on the self concept. (*c*) Modification of the ideal in a manner which resulted in greater acceptance of the pre-self as well as the post-self concepts. (*d*) Modification of the self in a manner which resulted in greater acceptance of the pre-ideal as well as the post-ideal concepts. (*e*) A firm reduction in the degree of suspiciousness of others, oversensitivity, and feelings of persecution. (*f*) A sound, healthy, internally and dynamically consistent pattern of psychological change which may be interpreted as meaning that psychological growth of a therapeutic nature occurred.

2. Doubling the number of group therapy sessions from 15 to 30 without increasing the total elapsed time in treatment results in significantly greater therapeutic movement.

3. Qualitative as well as quantitative differences in the pre-post change pattern results when the number of therapy sessions is increased from 15 to 30. These differences suggest that at 15 sessions the average group client was terminated just when he was on the threshold of his most serious therapeutic work.

4. The rate of therapeutic progress when considered as progression toward an end goal of a healthy configuration appears to gain momentum with successive increments of therapy although no single dimension appears to be a satisfactory measure of progress.

5. On the basis of mean group movements, there is considerable evidence for a common pattern of therapeutic movement which includes in its main outline an initial peaking which may best be described as a flight into health which occupies the first two weeks of therapy; a rapid and extensive regression and the initiation of the reintegrative phase in the second two weeks; and the rapidly increasing reintegration and consolidation of gain in the final two weeks.

6. The flight-into-health phenomenon occurs in those not participating in group

psychotherapy. Following the inevitable collapse, those not receiving group therapy show no indications of reintegration in a therapeutic sense but instead recover by re-erecting a structure only superficially different from the initial one.

7. Group psychotherapy actually seems to retard the development of the flight-into-health phenomenon and also dampens overall and short-term intra-individual self and ideal concept variability.

8. Group therapy results in progressive therapeutic psychological change while at the same time increasing stability of the self and ideal concepts by somehow introducing an internally consistent pattern with the implication of checks and balances to such change as does occur. That is to say, the movement seems sure and purposeful with far less "randomness" than occurs without therapy, and with far less actual variability.

9. Where short time limits are imposed on the term of therapy, intensive therapy results in maximum gain for the patient.

10. The effectiveness of group psychotherapy is not dependent on a self-recognized need for therapy before therapy begins. It is probable that group interaction results in recognition of the need after the group has begun.

11. Wherever possible, group-centered therapy should be an integral part of any serious program aimed at the treatment of alcoholics.

12. Further research in this area is needed.

REFERENCES

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.,* 1953, **17,** 327–333.

Brownfain, J. J. Stability of the self concept as a dimension of personality. *J. abnorm. soc. Psychol.,* 1952, **47,** 597–606.

Dymond, Rosalind. An adjustment score for *Q* sorts. *J. consult. Psychol.,* 1953, **17,** 339–342.

Dymond, Rosalind. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.,* 1955, **19,** 103–107.

Ends, E. J., & Page, C. W. A study of three types of group psychotherapy with hospitalized male inebriates. *Quart. J. Stud. Alcohol,* 1957, **18,** 263–277.

Fisher, R. *Statistical methods for research workers.* London: Oliver and Boyd, 1954.

Hartley, Margaret. Changes in the self concept during psychotherapy. Unpublished doctoral dissertation, Univer. of Chicago, 1951.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.

Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.

Mowrer, O. H. *Psychotherapy theory and research.* New York: Ronald Press, 1953.

Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.

Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

Rogers, C. R., & Dymond, Rosalind. *Psychotherapy and personality change.* Chicago: Univer. of Chicago, 1954.

Wirt, R. D. Further validation of the ego-strength scale. *J. consult. Psychol.,* 1955, **19,** 444.

(Accepted for publication August 25, 1958)

APPENDIX

Statements	Criterion Weights
1. I feel uneasy when talking with people.	(4)
2. I put on a false front.	(5)
3. I try to outdo or excel others.	(6)
4. I make strong demands on myself.	(6)
5. I often kick myself for the things I do.	(5)
6. I often feel humiliated.	(4)
7. I doubt my sexual powers.	(4)
8. I am much like the opposite sex.	(4)
9. I have a warm emotional relationship with others.	(9)
10. I am an aloof, reserved person.	(5)
11. I am responsible for my troubles.	(6)
12. I can be counted on to do the right thing.	(6)
13. I have a feeling of hopelessness.	(3)
14. I live largely by other people's values and standards.	(5)
15. I can accept most social values and standards.	(7)
16. I have few values and standards of my own.	(5)
17. I have a hard time controlling my sexual desires.	(4)
18. It's difficult for me to keep from criticizing and arguing with others.	(5)
19. Self-control is no problem for me.	(7)
20. I am often down in the dumps.	(4)
21. I am self-centered.	(5)
22. I usually like people.	(8)
23. I express my emotions freely.	(6)
24. Even when I am with people, I feel alone much of the time.	(4)
25. I feel like giving up trying to cope with the world.	(3)
26. I can live comfortably with the people around me.	(8)
27. My hardest battles are with myself.	(5)
28. I tend to be on my guard with people who are somewhat more friendly than I had expected.	(5)
29. I usually expect things to turn out all right.	(7)
30. I am stubborn.	(5)
31. I am critical of people.	(5)
32. I feel driven.	(5)
33. I am liked by most people who know me.	(6)
34. I have a feeling that I am not contributing enough to life.	(5)
35. I am sexually attractive.	(6)
36. I feel hopeless.	(2)
37. I can make up my mind and stick to it.	(6)
38. My decisions are not my own.	(5)
39. I often feel guilty.	(4)
40. I am a hostile person.	(3)
41. I am contented.	(6)
42. I am disorganized.	(3)
43. I don't really care about anything.	(3)
44. I feel comfortable and at ease in almost any social gathering.	(7)
45. I have to drive myself to get things done.	(5)
46. I often feel resentful.	(5)
47. I do things on the spur of the moment.	(4)
48. It's important for me to know what others think of me.	(5)
49. I don't trust my emotions.	(4)
50. It is pretty tough to be me.	(3)
51. I am a sensible person.	(7)
52. I'm just not facing things.	(4)
53. It doesn't bother me when others act, feel, or believe differently than I do.	(7)
54. I try to avoid thinking about my problems.	(4)
55. I have an attractive personality.	(6)
56. I am shy.	(5)
57. I need somebody to push me through on things.	(5)
58. I feel inferior.	(4)
59. I am no one. Nothing really seems to be me.	(2)
60. I am afraid of what other people think about me.	(5)
61. I am ambitious.	(6)
62. I despise myself.	(2)
63. I have initiative.	(7)
64. I shrink from facing a crisis or difficulty.	(5)
65. I just don't respect myself.	(4)
66. I usually run things in groups I'm in.	(5)
67. I take a positive attitude toward myself.	(8)
68. I stand up for my rights.	(6)
69. I am afraid of a full-fledged disagreement with a person.	(5)
70. I can't seem to make up my mind one way or another.	(4)
71. I am confused.	(3)
72. I am satisfied with myself.	(6)
73. I am a failure.	(1)
74. I am likable.	(6)
75. My personality is attractive to the opposite sex.	(6)
76. I am afraid of sex.	(4)
77. I have a horror of failure in anything I want to accomplish.	(5)
78. I feel relaxed and nothing really bothers me.	(6)
79. I am a hard worker.	(6)
80. I feel emotionally mature.	(8)
81. I am naturally nervous.	(3)
82. I really am disturbed.	(2)

Statements	Criterion Weights
83. All you have to do is just insist with me and I give in.	(4)
84. I feel insecure within myself.	(4)
85. I have to protect myself with excuses, with rationalizations.	(5)
86. I give in easily.	(5)
87. I am intelligent.	(7)
88. I feel superior.	(6)
89. I feel helpless.	(2)
90. I depend on myself.	(8)
91. I feel like arguing or telling people off.	(5)
92. I can't seem to let myself go and be my real self.	(5)
93. I am different from others.	(5)
94. I am unreliable.	(4)
95. I understand myself.	(9)
96. I am a good mixer.	(6)
97. I can handle most problems I am likely to face.	(7)
98. I am worthless.	(1)
99. I dislike my own sexuality.	(3)
100. I am not getting anywhere in life.	(5)

Vol. 73, No. 11

Whole No. 481, 1959

Psychological Monographs: General and Applied

PHOTOSENSITIZATION IN RELATION TO MEAN AND STANDARD DEVIATION VALUES[1]

RICHARD T. ZEGERS

Fordham University

THE RESEARCH described in this report covers a period of four years and was carried out as the result of a contract between the Office of Naval Research and Fordham University. This contract had its inception in a conversation between W. J. Crozier of Harvard University, William Berry of the Office of Naval Research, and the author. Crozier, making inferences from his own data, held that the human eye underwent a process of photosensitization when it operated at low levels of illumination. He believed that the evidence warranted an investigation to determine whether or not the process existed. The author agreed with him and undertook the work in 1952. The experimental design, to be described in Part 2, was explained to Crozier who felt it capable of revealing the process if it existed.

This process is defined as follows: If a monochromatic test signal light is flashed against a zero background and then against a series of monochromatic backgrounds, each of which is of a slightly higher intensity than the preceding, the energy in the flash of light, required for an *S* to see the flash 50% of the time, will decrease from that required at zero background, pass through a minimum at low levels of background intensity, and then rise according to the generally accepted principle of increase of ΔI with an increase of background intensity.

In order to test this hypothesis, four wavelengths of monochromatic light were selected. Each wavelength, when used as a signal light, was tested against itself and against the other three as backgrounds. In the choice of stimulus wavelengths, it was decided to use one from each of the main sections of the spectrum. A blue of 480 $m\mu$, a green of 509 $m\mu$, a yellow-green of 550 $m\mu$, and a red of 670 $m\mu$ were chosen. By using each one of these wavelengths, as a signal, against each as a background, the different color systems would be given an opportunity to reveal whether or not they functioned according to the demands of the hypothesis. After extensive testing of two

---

[1] The data of this report were gathered and processed under contract between the Office of Naval Research and Fordham University, contract Number N-onr-880-(00)-(01), covering the period from June 30, 1952 to July 1, 1956. The work was carried out by the Visual Laboratory of the Psychology Department under the direction of the author.

The author wishes to express his gratitude to the many people who assisted him in this research, especially to Donald J. Dillon, Amadeo P. Georgi, John A. Malone, George F. Nolan, Joseph E. Sturr, and Mrs. Rudolph L. Zlody who acted either as *S*s or as experimenters when the author, himself, was *S*.

Half of the data reported in Part 4 will constitute the experimental basis of the Ph.D. thesis of Mrs. Zlody. The data of Part 5 formed the basis for the M.A. theses of C. Claude Brodeur, Aime Hamann, William J. O'Halloran, and Jaime Bulatao. These theses were carried out under the mentorship of the author and were integrated by him into the research program of the visual laboratory.

My sincere appreciation is extended to E. Horace Siegler, Abraham Savisky, and Vincent Coates of the Perkin-Elmer Corporation for many technical and theoretical insights given during the construction of the colorimeter; to Philip Nolan and Bennett Sherman of the Farrand Optical Company for their assistance in developing the calibrating procedures; to Anne Anastasi for her discerning judgment in things statistical; to Forrest L. Dimmick for his helpful advice; and last but far from least to my former professor and mentor, Clarence C. Graham, who is always ready with the inspirational answers to a researcher's problems.

*S*s with these four wavelengths, the results were negative.

In order to explain these negative findings, the hypothesis was formulated that photosensitization is a phenomenon which manifests itself only as the result of the interaction of combinations of definite pairs of wavelengths. Once this hypothesis was suggested, a second hypothesis immediately presented itself. If photosensitization is a process which depends upon the interaction of specific signals with specific backgrounds, then the combinations of signal and background, which yield the phenomenon, depend upon the visual characteristics of the individual *S*. This secondary hypothesis was checked first. Crozier had offered evidence to show that the standard deviation of the logarithm of threshold energy varied periodically as a function of wavelength. It was decided to verify this pattern of variability and to use it as a guide for future experimentation. If individual *S*s, when tested, display the same periodic function with the maxima and minima located at exactly the same wavelengths, then it would be expected that photosensitization does not depend upon the peculiar characteristics of the *S*. Contrariwise, if it could be demonstrated that the maxima and minima of this periodic variability were definitely a function of the individual, then the negative results of the first part of the experiment might be explained in terms of the wavelengths which had been selected.

To investigate this latter hypothesis, a second experiment was devised. In this, two *S*s obtained frequency of seeing curves yielding means and standard deviations at 10 mμ intervals throughout the visual spectrum. An analysis of these data led to the tentative conclusion that maxima and minima of the standard deviations were functions of the individual involved. This necessitated a full-scale investigation of the problem of the periodic variability of the standard deviation. Moreover, it demanded investigation in a way that would give an index of the reliability of the data and show the scope of the phenomenon with reference to several psychophysical methods. The investigators agreed that if the standard deviation was really a measure of something over and above individual variability, and this to an important degree, evidences of such variability should be obtained regardless of the method used to gather the data.

When the data from this experiment failed to yield decisive results, specific investigation of the variability of an *S*'s standard deviation from session to session was undertaken. For this purpose, eight wavelengths, in two groups of four, were selected. Two groups of two *S*s each were tested on four of the eight wavelengths. This investigation consisted in obtaining 10 frequency-of-seeing curves per *S* per wavelength in an attempt to obtain a fair sample of day-to-day variability. Again the results were negative.

The final phase of the investigation consisted in studying the interaction of wavelength and field size on the pattern of the standard deviation. The mean values obtained from the frequency-of-seeing curves of this last phase were, because of calibration procedures, given in terms of radiometric units, i.e., microwatts. Calculations based on these data permitted presentation of the results in terms of a quanta-per-cone ratio.

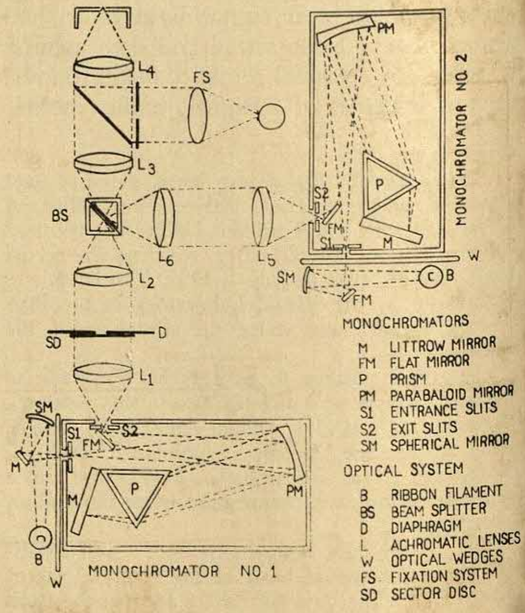


FIG. 1. Early model of the colorimeter.

APPARATUS

The apparatus grew in the course of the experiment. Parts 1 and 2 were carried out when the colorimeter had only two monochromators. By the time Parts 3–6 were begun, two more monochromators had been added. With the initiation of Part 7, a thermocouple, amplifier, Ballantine voltmeter, powerpack, and auxiliary equipment had been acquired, making it possible to estimate the number of quanta involved in the threshold measurements. Schematic diagrams of the earlier and present apparatus are presented in Fig. 1 and 2, respectively. Basically, the earlier system would be the same as the present one when only Monochromators 2 and 4 are used. A detailed description of the present apparatus will be simplified if the monochromators, the optical system, and finally the auxiliary equipment required for its operation are considered separately.

Monochromators

The four monochromators are Perkin-Elmer Universal Monochromators, Model 83. These are Littrow-type prism monochromators. In this research, each monochromator was equipped with a 6-v., 9-amp. tungsten vertical ribbon filament bulb,

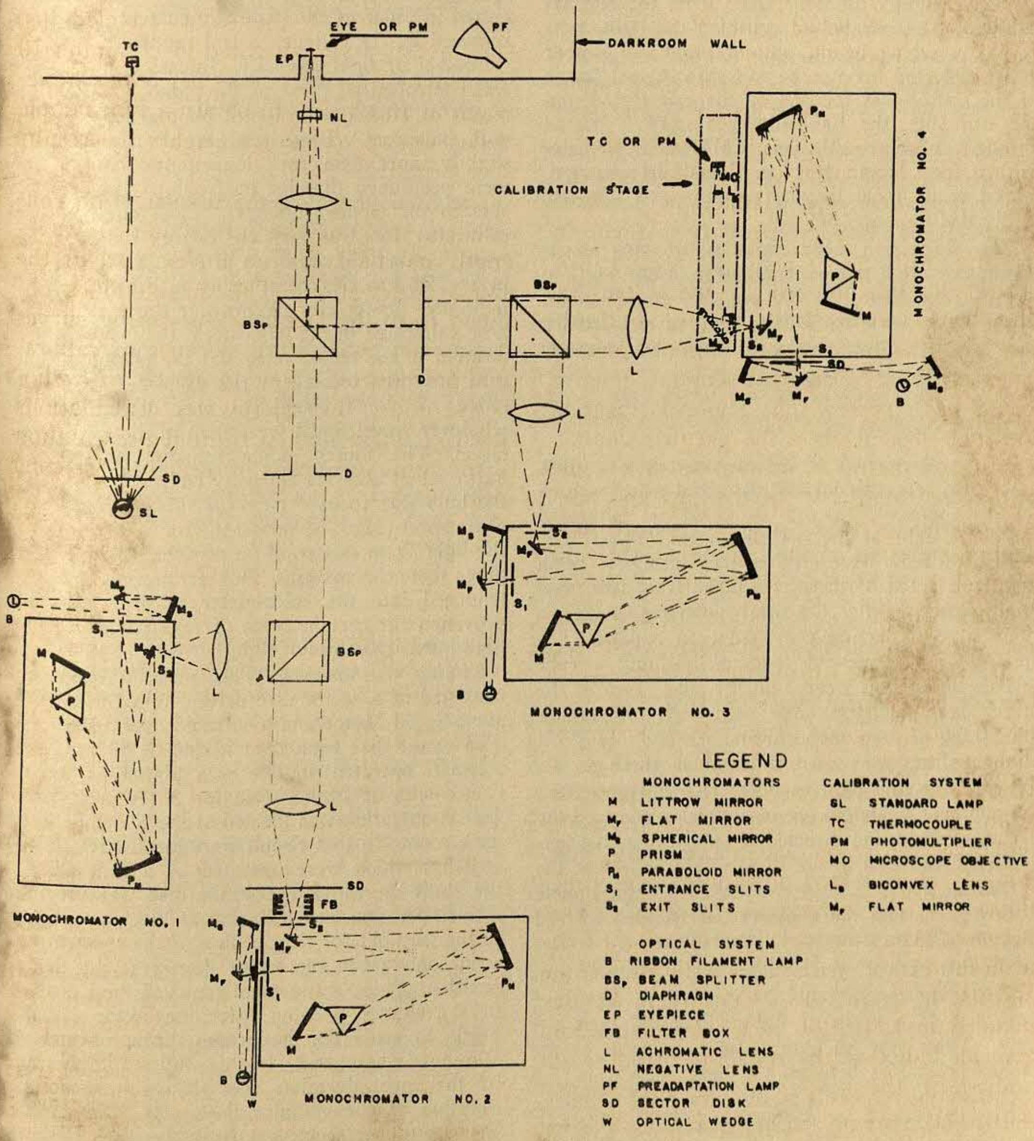


FIG. 2. The Fordham colorimeter.

as a light source. The diverging light from the filament was collected and reflected by a spherical mirror onto a flat mirror which deflected the rays of light through an intensity control system, to be described later, and located a real image of the filament on the entrance slits of the monochromator. These slits were controlled by a micrometer, calibrated in millimeters. Slit widths could be reproduced to .001 of a millimeter. The light passing through the slits became a diverging beam which was collected and collimated by a paraboloid mirror. The light reflected from this mirror then passed through the prism, fell on the Littrow mirror and was reflected back through the prism to the paraboloid. The return path of the light was at a slightly different angle from the original pathway. The paraboloid reflected the light, now in a converging beam, onto a small flat mirror which reflected it onto the exit slits. A real image of the source was focused on the exit slits. From the exit slits, the light diverged until it passed through an achromatic lens, 1.875 in. in diameter with a focal length of 11 in. This lens was positioned at its focal length from the exit slits and collimated the beam of light.

The dispersion characteristics of the monochromator are 5 mμ per millimeter of slit width at 400 mμ and 30 mμ at 700 mμ. To guard against stray light, various Kodak Wratten or Corning colored glass filters were used across the beams at the exit slits.

Optical System

In the description of the monochromators given above, an arrangement was described whereby each of the monochromators delivered a collimated beam of light to the optical system. These four beams of light were combined by means of beam splitters. Each of these was made from two glass right angle prisms. On the hypotenuse surface, a half silver reflecting coat had been applied. The prisms were cemented together, yielding a cube whose dimensions were 3 in. per side. Two of the three beam splitters were used each to combine the light of two monochromators and the third beam splitter was used to combine these double beams. The four light beams were then transmitted to another double convex lens which converged the beam of light and directed it onto a negative lens which formed an image of the light source at the artificial pupil. The combination of the double convex lens and the negative lens yielded a focal length of 15 in.

In this optical system, there were two places at which diaphragms could be inserted to determine the size of a beam of light which fell upon the eye. There was an optical pupil for each of the monochromators between the light source and the spherical mirror. This position was used whenever it was necessary to diaphragm each of the four beams separately. Where it was permissible to diaphragm two beams simultaneously, the focal points marked on Fig. 2 were employed.

Accessory Equipment

Power supply. The power for the light sources was delivered by a Sorenson Nobatron, Model E-6-40A. This instrument received power from the ordinary 110 AC house mains. It transformed this current down to six or seven volts, dependin on the operation, and converted, by means o selenium rectifiers, the alternating current to direc current. An electronic voltage control apparatus regulated the current and kept it constant within the limits of one tenth of one per cent. A 6-v. storage battery was connected across the Nobatron output in order to eliminate any current ripples not removed by the voltage control regulator.

Control of light intensity. Intensity of the light was controlled on three of the monochromators by means of Kodak neutral tint photographic wedges with balancers. These wedges have a range of slightly more than four logarithmic units. They were positioned directly in front of the entrance slits to the monochromators. Each one had been calibrated for transmission factors at the Electrical Testing Laboratories in New York City and in the laboratories here, using a Farrand photomultiplier photometer. Whenever the wedges have been calibrated in the laboratory, they have been in position in the colorimeter and the photomultiplier tube has been located at the eyepiece. The logarithm of the transmission was determined for whatever wavelengths were used in the experimentation. The fourth monochromator employed an extra set of slits for intensity control. Light from the bulb was focused on these slits by means of a short focal-length, spherical mirror and the path of the light from this point on was the same as in th other monochromators. This arrangement was introduced into the colorimeter in order to make provision for certain types of research which need high intensity values of light.

Timing mechanisms. This piece of apparatus consisted of a sector disc driven by a Bodine Constant Speed Motor and a solenoid operated shutter. The sector disc rotated at a rate of 50 rpm. The solenoid operated shutter was placed at the entrance slits of Monochromator 2. A push button for its operation was located at the *S*'s chair in the dark room. In this circuit were located two microswitches which were operated by a cam fixed to the shaft on which the sector disc was mounted. Their operation was such that an electrical impulse to the solenoid for the opening of the shutter could occur only when the sector disc was in a definite position. This arrangement removed the possibility of the shutter opening after the sector disc had begun to cross the diaphragm through which th stimulus light passed. In this way a stimulus ligh of the same duration was always insured. Th duration of the stimulus flash was a function o the size of the sector cut from the disc. Differen stimulus durations were obtained by using differe sector discs. In the present experiment, the stimulus duration was varied from one part to another and will be stated where appropriate.

Monitoring. In order to ascertain the constancy of the power applied to the light source, a voltmeter was connected across the storage battery, and an ammeter was connected in series with the light source. These were checked regularly throughout the testing sessions. As a further intensity control, a Weston photronic cell was placed in close proximity to whichever one of the four light sources required critical observation. The cell was connected in series with a microammeter. The deflections of this latter instrument were watched for any indication of shifts in the energy emitted by the light source during the progress of the research.

Calibration: Equipment and Procedures

Wavelength

Wavelength calibration was carried out by using a 100-w. AH4 General Electric mercury arc and Geissler tubes filled with neon, hydrogen, argon, and helium. The arc or a Geissler tube was substituted for the light source in the monochromator which was to be calibrated. The wavelength drum was then turned until one of the identifiable spectral lines was focused in the eyepiece and the drum setting was recorded. The drum was then moved until another line was picked up. A group of some 15 lines spread across the spectrum was located. A graph was then plotted of wavelength versus drum settings. This plot yielded a smooth curve. Other spectral locations were then obtained from the graph.

Intensity

A Farrand photomultiplier photometer was used for purposes of intensity calibration. The photometer head was furnished with one of three phototubes depending upon the type of calibration required. The phototubes were R.C.A. 1P21, 1P22, and 6217. All tubes were calibrated for relative spectral sensitivity for each wavelength employed during this research. For the greater part of the time radiometric measurements in arbitrary units were sufficient. The units employed were the scale units of the galvanometer of the photometer corrected to give equal response regardless of wavelength.

After the laboratory acquired the thermocouple with its amplifier and indicating equipment (a Ballantine voltmeter), the units in which the apparatus was calibrated were microwatts from which calculations could be made to the number of quanta involved. The method whereby this was accomplished was to calibrate the thermocouple against a carbon filament standard lamp. This lamp was calibrated by the Electrical Testing Company for flux output in terms of microwatts per centimeter square. The lamp was calibrated at four different amperage values. A plot of voltage readings from the Ballantine versus the calibration values of the standard lamp yielded a straight line which passed through the calibration points and the origin. The slope of this line was the conversion factor whereby the voltage readings of the Ballantine were converted into microwatts.

While, ideally speaking, it would have been desirable to place the thermocouple at the eyepiece and make readings directly from the Ballantine, there was not enough light at the shorter wavelengths to permit this. Consequently, an indirect method was employed. A special bracket was designed which, by means of lenses and an inverted microscopic objective, focused the output of Monochromator 4 on the thermocouple. This was so mounted that it could be precisely manipulated in each of the three main axes. With this mounting it was possible to locate the thermocouple so that it received all the light from the monochromator. The monochromator was calibrated for whatever combination of wavelength and pass band was desired. After this the thermocouple was removed and the phototube placed so that its center received the light which formerly fell on the thermocouple. The scale unit deflections of the phototube galvanometer were obtained and a ratio of microwatts per scale unit deflection was calculated. When the phototube was employed to calibrate the colorimeter, it was placed at the artificial pupil at the point where the *S* located his eye.

PROCEDURE

This research treats of seven related topics. Part 1 deals with the photopic luminosity curves of the four *S*s who carried out the major portion of this experimentation. This is included to demonstrate that these *S*s had essentially normal color vision. Part 2 reports the data obtained from the investigation of the problem of photosensitization, i.e., the differential threshold for foveal vision at low background levels. Part 3 was a preliminary investigation into the variability of the standard deviation with wavelength. The findings obtained here led directly to Part 4 which consists in an extensive study of the same problem. Parts 5 and 6 were logical developments of Part 4. Part 5 studied the variation of the standard deviation under fixed conditions of *S*, field size, and method, at a few specified wavelengths. This yielded an estimate of the minimum variability of the standard deviation over a period of time. Part 6 treats of the same problem with the added parameter of field size, introducing the possibility of greater variability. The last part presents quantum estimates of the mean values of the data from the preceding part. Because the precise experimental conditions varied considerably from one part to another, the part procedures will be described each in its own section of the results.

SUBJECTS

Nine *S*s took part in this investigation; eight males and one female. The age ranged from the middle forties for one *S*, the middle thirties for another, to the mid-twenties for the remaining *S*s.

RESULTS

Part 1: Photopic Luminosity Curves of Four Ss

The photopic luminosity curves of the four *S*s whose data will form the substance of Parts 2, 3, and 4 of this monograph are presented here. The field used was a circular disc subtending one degree of visual angle. It flashed into the center of the fovea for durations of one-twentieth of a second. The data for *S*s I and II are taken directly from the work of Dillon (1954) because of the identity of the *S*s. Each value given for these two *S*s represents the average of 50 determinations. The ascending method of limits was employed. Five threshold determinations were made at each of 31 wavelengths per session, and there were 10 sessions. The order of presentation of the wavelengths was varied systematically from session to session. The curves of *S*s III and IV were derived from frequency-of-seeing curves. One frequency-of-seeing curve was obtained for each wavelength. Each curve was based upon 50 determinations made at each of five energy levels. The points were plotted on probit paper with the logarithm of the energy as the abcissa and percentage of seeing as the ordinate. The data fell on straight lines and the fit was made by eye. The 50% seeing point was taken as the mean value. It had been planned to report luminosity data derived from the data of Part 4 of the work, but noise developing in the phototube made some of the mean values suspect and it was decided not to include them in this paper.

At the time when the curves were obtained, the laboratory had not been equipped with the thermocouple and its auxiliary apparatus. The units of energy used were the scale deflections of the photometer galvanometer corrected for the differential sensitivity of the 1P22 phototube. After the new equipment had been put into working condition, *S*s I and II made extensive judgments at wavelength 550 mμ and the microwatt values for these data were obtained. It is worthy of note that the wedge settings were within .02 log units of those obtained during the original data gathering. The differences among the wedge settings of the four *S*s at 550 mμ were calculated and, by means of these differences, the microwatt

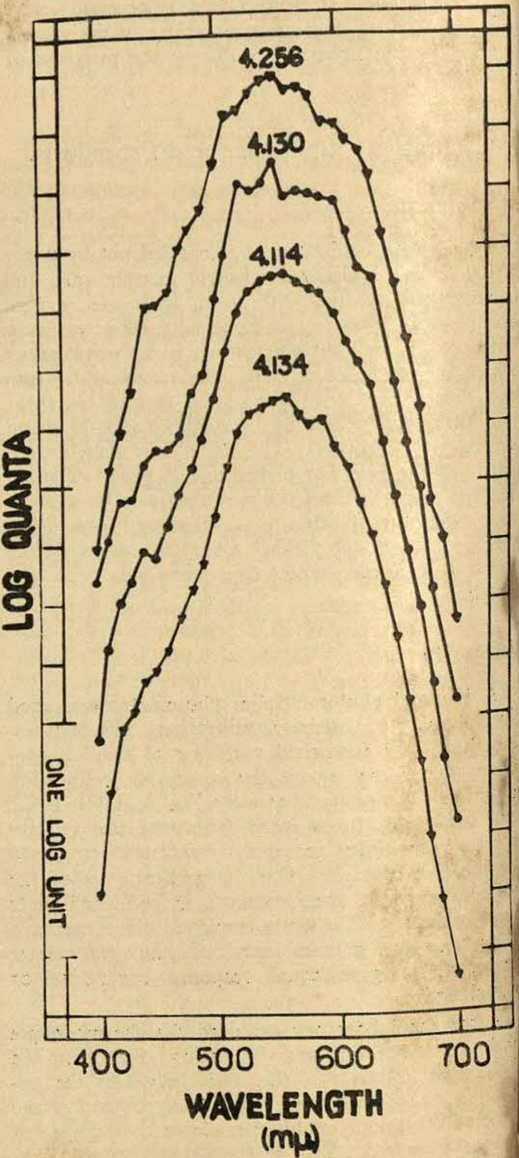


FIG. 3. Foveal luminosity curves for four *S*s. Number indicates log quanta values for each *S* at $\lambda = 550$ mμ.

values for *S*s III and IV were obtained. The 550 mμ values were then used to calculate the microwatt values for the four *S*s at all wavelengths. The microwatt values were then converted into ergs and divided by the proper quantal value (hν) for each wavelength.

The data of the luminosity curves for the four *S*s are given in Table 1 and Fig. 3. The number at the maximum of each curve is the log quanta threshold energy value for wavelength 550 mμ. The ordinate scale is in .10 log units. The curves of all four *S*s depart from the smoothness of the I.C.I. curve. However, in units of microwatts they have been compared with those recently published by Hsia and Graham (1952), and Hurvich and Jameson (1953), and so closely agree with them that any discrepancies are felt to be accounted for in terms of individual differences. All show inflections varying in amount between 440 and 450 mμ, and again at 490 mμ. These inflections have been interpreted by Wald (1948) to represent absorption of light energy by macular pigment at these wavelengths. Others have claimed that the same irregularities are due to the activity of the different color receptor systems within the eye. Except for these small inflections of the curves there is apparently no other significant deviation from the smooth curve. All of the *S*s show the maximum at 550 mμ, excepting *S* I who shows it at 560 mμ.

Comparing the threshold values between the point of maximum sensitivity and 400 mμ, we find the following: *S* I exhibits the greatest range, approximately 2.26 log units;

TABLE 1

LOGARITHM OF THE NUMBER OF QUANTA REQUIRED FOR THRESHOLD FOVEAL LUMINOSITY CURVES FOR FOUR *S*s

Wavelength (mμ)	*S* I	*S* II	*S* III	*S* IV
400	6.228	6.061	5.9435	6.2693
410	5.801	5.692	5.7212	5.9307
420	5.512	5.486	5.5603	5.7812
430	5.450	5.405	5.5134	5.5733
440	5.322	5.281	5.3996	5.2317
450	5.274	5.301	5.3413	5.2378
460	5.196	5.161	5.3432	5.1961
470	5.049	5.017	5.2965	4.9848
480	4.918	4.908	5.0982	4.8860
490	4.829	4.798	5.0307	4.8155
500	4.617	4.625	4.6923	4.6428
510	4.408	4.420	4.5106	4.4449
520	4.287	4.262	4.2125	4.3997
530	4.199	4.179	4.2473	4.3405
540	4.161	4.130	4.2210	4.2902
550	4.134	4.114	4.1300	4.2560
560	4.111	4.114	4.2819	4.3385
570	4.207	4.146	4.2512	4.3266
580	4.223	4.161	4.2594	4.3524
590	4.212	4.196	4.2791	4.4693
600	4.342	4.272	4.3128	4.4581
610	4.418	4.401	4.4323	4.5384
620	4.581	4.476	4.5804	4.5940
630	4.716	4.573	4.6181	4.6883
640	4.928	4.812	4.8198	4.9433
650	5.146	5.053	5.0688	5.1506
660	5.420	5.290	5.3205	5.3875
670	5.697	5.526	5.5178	5.6700
680	6.004	5.848	5.9567	6.0592
690	6.265	6.164	—	6.2389
700	6.576	6.441	6.4211	6.5755

S II, a range of 2.08; *S* IV, a range of 1.91; and *S* III, a range of 1.72. These ranges are considerably less than those which were set up by the I.C.I. photopic luminosity curve and their range corresponds approximately to the respective ages of the *S*s.

Part 2: The Differential Threshold in Foveal Vision, i.e., Photosensitization

The hypothesis of photosensitization, as previously mentioned, states that whenever a signal light is made to appear against graded series of background intensities ranging from zero value to values well above the cone threshold, the energy required by the flash of light in order to be seen one-half of the time is minimal, not at zero background, as would be expected, but at some low background value. In other words, the presence of a dim surround of monochromatic light should lower the threshold value of a signal light seen against the background.

For this part of the experiment, four monochromatic bands of light, 5 mμ in width, centered at 480 mμ, 509 mμ, 550 mμ, and 670 mμ, respectively, were employed. These four wavelengths were used in the 16 possible combinations as background and signal. For the background, a range of 9–12 intensity levels was used. In all cases, the lowest level was that of zero intensity. The second intensity level was approximately at rod threshold for the dark adapted eye and the following levels were arranged in steps, each of which was half a logarithmic unit above that of the preceding step. In this way, the range of background from none at all, through rod threshold, up to and beyond cone threshold was investigated. The highest background in all cases represented the maximal amount of light which the apparatus was capable of delivering under the prescribed conditions of wavelength and pass band. The method of determining the threshold value for the signal light was that of constant stimuli. Five points within the range of the *S*'s doubtful responses were roughly determined. At each of these five determined points, the *S* was given, in completely random fashion, 50 stimulations to which he reported seeing or not seeing the flash of light. Hence, each frequency-of-seeing curve is based upon 250 responses. For ease in manipulating the apparatus, each of these five points was separated by approximately .20 to .25 of a logarithmic unit. The frequency-of-seeing curves were plotted on probit paper, the abcissa being the logarithm of the intensity and the ordinate the percentage of response. The curves describing the data were all straight lines and from these lines the mean and standard deviations were taken. The mean value was calculated from the 50% point and the standard deviation from

$$\frac{\text{Log } E_{(.8413)} - \text{Log } E_{(.1586)}}{2}$$

Two *S*s were employed in this part of the experiment. The signal light was fixated in the fovea by means of a tiny red fixation point. In the early stages of the experimentation, the signal light subtended a visual angle of 1.5°. After the experiment was in progress, the field of the signal light was reduced to one degree of visual angle, to better insure the fact of purely foveal stimulation. The signal light was centered in the background which subtended a visual angle of approximately 9.5°.

The basic data for this part of the experiment are contained in Table 2 and Fig. 4. In each of these experimental situations, the first step in the procedure was to obtain a mean and a standard deviation from a frequency-of-seeing curve for the signal light under conditions of zero background. The 1.5° field was used by *S* I for the four wavelengths of the signal light against the 480 mμ background light. *S* II also used it for the 480 mμ signal light against the 480 mμ background. At this point, the field size was reduced to one degree. This change has the effect of slightly raising the threshold. Beyond this, the change of the field has no real effect upon the experiment. Each signal light was tested four times with a zero background in the case of each *S*. Due to the change in field size, *S* I obtained only three absolute threshold values under

TABLE 2

MEAN AND STANDARD DEVIATION IN ARBITRARY UNITS OF THE LOGARITHM OF THE ENERGY FOR TWO *S*s AND FOUR TEST LIGHTS AGAINST FOUR BACKGROUNDS AT VARIOUS LEVELS OF BACKGROUND ILLUMINATION

S	Logarithm of Background Intensity in Arbitrary Units	Background Light							
		480 mμ		509 mμ		550 mμ		670 mμ	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Signal Light 480 mμ									
I	0.0	−0.397	0.133	−0.438	0.180	−0.350	0.167	−0.155	0.120
	1.0	—	—	−0.505	0.135	−0.467	0.120	—	—
	1.5	—	—	−0.445	0.166	−0.467	0.151	—	—
	2.0	—	—	−0.474	0.130	−0.467	0.144	—	—
	2.5	−0.387	0.165	−0.421	0.132	−0.503	0.144	−0.288	0.120
	3.0	−0.328	0.127	−0.433	0.144	−0.432	0.132	−0.216	0.131
	3.5	−0.343	0.192	−0.447	0.151	−0.432	0.132	−0.337	0.153
	4.0	−0.466	0.148	−0.457	0.147	−0.479	0.120	−0.385	0.142
	4.5	−0.237	0.102	−0.450	0.109	−0.360	0.108	−0.385	0.110
	5.0	−0.304	0.108	−0.428	0.120	−0.384	0.120	−0.288	0.142
	5.5	−0.286	0.199	−0.306	0.139	−0.168	0.132	−0.404	0.153
	6.0	−0.122	0.204	−0.101	0.170	−0.071	0.108	−0.312	0.164
	6.5	—	—	—	—	—	—	−0.288	0.164
	7.0	—	—	—	—	—	—	−0.312	0.175
	7.5	—	—	—	—	—	—	−0.070	0.153
II	0.0	−0.229	0.144	*−0.148*	0.133	*−0.238*	0.138	*−0.251*	0.169
	1.0	—	—	−0.162	0.157	−0.045	0.145	—	—
	1.5	−0.157	0.149	−0.167	0.169	−0.115	0.120	—	—
	2.0	−0.169	0.127	−0.109	0.133	−0.197	0.138	—	—
	2.5	−0.229	0.163	−0.131	0.133	−0.228	0.160	—	—
	3.0	−0.152	0.142	−0.119	0.145	−0.223	0.172	—	—
	3.5	−0.229	0.135	−0.131	0.145	−0.146	0.124	−0.276	0.125
	4.0	−0.272	0.168	−0.114	0.121	−0.243	0.182	−0.148	0.120
	4.5	−0.296	0.125	−0.041	0.145	−0.211	0.165	−0.208	0.145
	5.0	−0.217	0.204	−0.072	0.085	−0.024	0.148	−0.280	0.142
	5.5	−0.253	0.168	0.161	0.109	0.123	0.160	−0.220	0.145
	6.0	−0.161	0.113	0.353	0.121	0.281	0.165	−0.078	0.169
	6.5	—	—	—	—	—	—	−0.022	0.120
	7.0	—	—	—	—	—	—	−0.122	0.108
	7.5	—	—	—	—	—	—	0.351	0.140
Signal Light 509 mμ									
I	0.0	*−1.103*	0.136	*−0.739*	0.168	*−0.841*	0.132	*−0.818*	0.110
	1.0	—	—	−0.836	0.127	−0.903	0.125	—	—
	1.5	−1.177	0.143	−0.808	0.115	−0.846	0.137	—	—
	2.0	−1.179	0.134	−0.801	0.138	−0.834	0.125	—	—
	2.5	−1.200	0.138	−0.854	0.155	−0.868	0.148	—	—
	3.0	−1.211	0.132	−0.824	0.142	−0.784	0.137	—	—
	3.5	−1.234	0.205	−0.815	0.152	−0.823	0.114	−0.894	0.110
	4.0	−1.234	0.151	−0.780	0.157	−0.823	0.114	−0.927	0.099
	4.5	−1.204	0.109	−0.819	0.150	−0.789	0.125	−0.807	0.176
	5.0	−1.149	0.131	−0.808	0.150	−0.743	0.137	−0.796	0.154
	5.5	−1.149	0.161	−0.635	0.170	−0.573	0.148	−0.840	0.110
	6.0	−1.025	0.151	−0.301	0.108	−0.288	0.125	−0.905	0.143
	6.5	—	—	—	—	—	—	−0.840	0.165
	7.0	—	—	—	—	—	—	−0.686	0.143
	7.5	—	—	—	—	—	—	−0.466	0.176

TABLE 2 (continued)

II	0.0	*−0.754*	0.184	*−0.631*	0.117	*−0.701*	0.135	*−0.628*	0.115
	1.0	—	—	−0.631	0.117	−0.605	0.104	—	—
	1.5	−0.773	0.156	−0.596	0.140	−0.763	0.140	—	—
	2.0	−0.800	0.161	−0.608	0.128	−0.616	0.124	—	—
	2.5	−0.796	0.128	−0.631	0.163	−0.644	0.145	—	—
	3.0	−0.800	0.150	−0.608	0.117	−0.568	0.101	—	—
	3.5	−0.796	0.161	−0.596	0.082	−0.586	0.115	−0.559	0.131
	4.0	−0.823	0.115	−0.585	0.186	−0.547	0.143	−0.531	0.127
	4.5	−0.800	0.150	−0.608	0.163	−0.556	0.094	−0.711	0.152
	5.0	−0.773	0.138	−0.596	0.105	−0.352	0.159	−0.692	0.127
	5.5	−0.573	0.131	−0.491	0.117	−0.094	0.131	−0.605	0.115
	6.0	−0.356	0.104	−0.352	0.093	—	—	−0.547	0.110
	6.5	—	—	—	—	—	—	−0.508	0.150
	7.0	—	—	—	—	—	—	−0.301	0.110
	7.5	—	—	—	—	—	—	−0.174	0.115
Signal Light 550 mμ									
I	0.0	*−1.309*	0.113	*−1.149*	0.165	*−1.050*	0.132	*−1.106*	0.173
	1.0	—	—	−1.275	0.132	−1.006	0.165	—	—
	1.5	−1.287	0.171	−1.149	0.198	−1.149	0.143	—	—
	2.0	−1.236	0.160	−1.127	0.176	−1.105	0.187	—	—
	2.5	−1.331	0.170	−1.149	0.132	−1.050	0.132	—	—
	3.0	−1.251	0.182	−1.204	0.143	−1.050	0.143	—	—
	3.5	−1.214	0.186	−1.138	0.121	−1.050	0.143	−1.041	0.152
	4.0	−1.203	0.190	−1.193	0.154	−0.973	0.153	−1.073	0.206
	4.5	−1.300	0.164	−1.171	0.132	−0.973	0.132	−1.062	0.130
	5.0	−1.236	0.195	−1.105	0.154	−0.973	0.132	−1.041	0.065
	5.5	−1.192	0.173	−0.951	0.132	−0.929	0.132	−1.149	0.173
	6.0	−1.103	0.244	−0.545	0.121	−0.622	0.153	−1.041	0.087
	6.5	—	—	—	—	—	—	−1.030	0.130
	7.0	—	—	—	—	—	—	−0.813	0.173
	7.5	—	—	—	—	—	—	−0.629	0.108
II	0.0	*−1.220*	0.115	*−0.995*	0.112	*−1.224*	0.100	*−0.957*	0.113
	1.0	—	—	−0.966	0.101	−1.190	0.245	—	—
	1.5	−1.225	0.200	−0.977	0.123	−1.246	0.134	—	—
	2.0	−1.203	0.175	−0.981	0.146	−1.135	0.134	—	—
	2.5	−1.125	0.177	−0.954	0.184	−1.202	0.134	—	—
	3.0	−1.092	0.140	−0.894	0.218	−1.079	0.111	—	—
	3.5	−1.103	0.155	−0.977	0.166	−1.035	0.111	−0.954	0.144
	4.0	−1.072	0.160	−0.966	0.123	−0.957	0.145	−0.957	0.113
	4.5	−1.082	0.188	−0.954	0.146	−0.968	0.189	−0.932	0.133
	5.0	−0.963	0.153	−0.965	0.146	−0.957	0.156	−0.972	0.122
	5.5	−0.977	0.164	−0.802	0.135	−0.812	0.156	−0.855	0.127
	6.0	−0.852	0.111	−0.707	0.164	−0.456	0.111	−0.855	0.138
	6.5	—	—	—	—	—	—	−0.695	0.138
	7.0	—	—	—	—	—	—	−0.484	0.173
	7.5	—	—	—	—	—	—	−0.262	0.093
Signal Light 670 mμ									
I	0.0	*−0.673*	0.113	*0.300*	0.114	*0.378*	0.153	*0.244*	0.099
	1.0	—	—	0.379	0.148	0.378	0.194	—	—
	1.5	—	—	0.368	0.148	0.344	0.166	—	—
	2.0	—	—	0.368	0.148	0.355	0.180	—	—
	2.5	−0.717	0.120	0.436	0.136	0.469	0.166	—	—
	3.0	−0.789	0.156	0.391	0.136	0.469	0.153	—	—
	3.5	−0.831	0.184	0.266	0.136	0.378	0.166	0.189	0.155
	4.0	−0.774	0.134	0.334	0.148	0.469	0.153	0.322	0.110
	4.5	−0.750	0.114	0.300	0.159	0.458	0.125	0.322	0.141
	5.0	−0.768	0.103	0.368	0.125	0.424	0.139	0.288	0.099
	5.5	−0.731	0.127	0.459	0.136	0.537	0.166	0.260	0.128
	6.0	−0.623	0.175	0.641	0.114	0.674	0.208	0.344	0.152
	6.5	—	—	—	—	—	—	0.443	0.139
	7.0	—	—	—	—	—	—	0.553	0.143
	7.5	—	—	—	—	—	—	0.939	0.137

TABLE 2 (continued)

II	0.0	*0.300*	0.107	*0.503*	0.139	*0.235*	0.134	*0.322*	0.112
	1.0	—	—	0.449	0.137	0.226	0.116	—	—
	1.5	0.275	0.093	0.449	0.086	0.190	0.108	—	—
	2.0	0.318	0.114	0.469	0.148	0.204	0.094	—	—
	2.5	0.300	0.091	0.424	0.109	0.231	0.101	—	—
	3.0	0.350	0.102	0.478	0.107	0.276	0.090	—	—
	3.5	0.352	0.095	0.401	0.137	0.287	0.103	0.275	0.114
	4.0	0.363	0.095	0.446	0.137	0.249	0.112	0.237	0.085
	4.5	0.350	0.109	0.394	0.123	0.349	0.067	0.260	0.123
	5.0	0.316	0.114	0.471	0.125	0.334	0.105	0.237	0.157
	5.5	0.504	0.118	0.435	0.196	0.506	0.090	0.271	0.134
	6.0	0.686	0.068	0.585	0.141	0.614	0.112	0.376	0.103
	6.5	—	—	—	—	—	—	0.580	0.143
	7.0	—	—	—	—	—	—	0.562	0.123
	7.5	—	—	—	—	—	—	0.712	0.157

the conditions of a one degree field. *S* II obtained four absolute threshold values for each signal light for all of the wavelengths except the 480 mμ signal light against the 480 mμ background. These absolute values, when compared with each other, serve as an index of the consistency of the *S*s throughout the period of this experimentation. The mean and standard deviations obtained under these conditions of zero background have been printed in italics in Table 2 for ease of comparison.

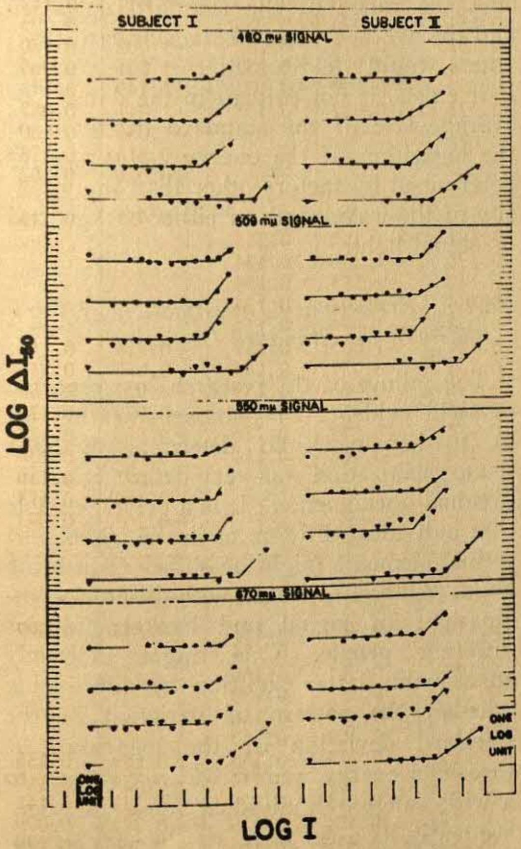


FIG. 4. Log ΔI.60 values for various log I0 values. Background: Open circles —480 mμ, closed circles —509 mμ, open triangles —550 mμ, closed triangles —670 mμ.

As shown by these figures, the shift in mean absolute threshold values throughout a time interval of 24 months is less than three-tenths of a logarithmic unit. In five of the eight cases it is less than .15 of a logarithmic unit. Denton and Pirenne (1954) accept as permissible a shift in mean absolute threshold values of .20 of a logarithmic unit, under conditions of maximum stability. Hecht (1944) presents data showing a variability of .30 of a logarithmic unit in absolute threshold determinations. It is clear that the variability of the *S*s of this experiment lies within established limits.

An examination of the data contained in Table 2, background by background, and *S* by *S*, shows that for all background and signal light combinations (except for the three highest background intensities under all these conditions), the shift in mean values of the intensity of the signal is less than .29 of a logarithmic unit. Out of the 32 possible combinations of *S*, signal and background lights, we find a range of mean values of signal light intensity of less than .15 log units in 18 situations, of less than .20 in nine others, of less than .25 in four others, and in only one case does the value exceed .25. In only one combination of signal and background lights (480 mμ, signal; 670 mμ, background) do both *S*s show a range of over .20 log units in mean variability of the signal light at the lower levels of background intensity. In all 32 situations, the two or three highest background intensities show a mean value of the signal light whose magnitude is definitely increasing. This region of background intensities is interpreted as that background level at which the value of the differential threshold begins to increase. If the data pertaining to the three highest levels of background intensity are not considered for the moment, the following ranges of mean scores have been obtained. In approximately three-fourths of the *S*, test-light, background combinations, the range of mean scores is less than .20 of a logarithmic unit. The remaining cases, where the range is more than this value, show the variability of one *S* to be large while that of the other is small. There is only a single case where both *S*s show a range of mean variability of more than .20 of a logarithmic unit. In this isolated case there is no regular pattern in the variation of the mean. It is difficult to conclude to a process of photosensitization when the mean variability at different background intensities is not larger than that of the variability of the absolute threshold.

Weber's law has been shown not to hold throughout the range of intensities to which the eye is capable of responding (Hecht, 1934). The commonly accepted opinion is that the absolute value of $\triangle$I increases with an increase in I. The evidence of the present research is that under conditions where the $\triangle$I flash is completely confined to the fovea its value is that of the absolute threshold for all levels of the background intensity which are below the cone threshold itself.

The simplest explanation for the constancy of $\triangle$I through this range is that the rate of breakdown of photochemical substances is so slow and the amount so small that the restoration process is easily able to counterbalance it and to leave the cones in a state in which they possess a concentration of photochemical substances which is very close to 100%. The rise in $\triangle$I at those levels of background illumination which are above the cone threshold would then be explained by the fact that the rate of restoration is not able to bring the concentration of photochemical substances back to a level closely approximating the maximum.

A study of the standard deviations given in Table 2 shows that they vary in an unpredictable fashion from session to session within a wavelength over a range of .03 to .145 log units. It is felt that because of the irregularities manifested in the standard deviations, they represent, purely and simply, day-to-day variability. If the size of the standard deviation were being determined by variations in the wavelength of either the signal or the background, or their intensity, or both, one would expect that both the range of the standard deviation and the average value of the standard deviations would yield a pattern. The pattern, if it exists, is not evident in the data. The average size of the standard deviation of the logarithm of the energy appears to be determined by factors other than the intensity of the wavelength of either background or signal light.

Part 3: Preliminary Investigation of Variability of the Standard Deviation

The failure of the research just reported to yield evidence of photosensitization led to further work to determine whether photosensitization was very definitely an individual phenomenon. It is a priori possible that individual differences in the makeup of retinal elements might be sufficient to bring about photosensitization with different combinations of signal and background for different people. One might indirectly answer this latter question by discovering whether the pattern of variation of the standard deviation of the logarithm of threshold energy was or was not subject to a wide range of individual variation. Consequently, two *S*s were set the task of obtaining frequency-of-seeing curves at wavelengths throughout the visual spectrum from 410 to 700 mμ. The wavelengths tested were separated by intervals of 10 mμ and the pass band was 5 mμ. At each of these points, each of the two *S*s obtained a frequency-of-seeing curve according to the same method as was employed in Part 2, namely, five points were chosen within the range of doubtful response and the *S* was tested with 50 determinations at each of these five points. As previously, the intensity of the stimuli was varied in a random fashion. The signal light subtended a visual angle of one degree of arc. The fixation point was a single red dot centered on the field where the signal light would flash. The results of this experiment are given in Table 3 and Fig. 5. The curves of Fig. 5 lend superficial support to the concept that the maxima and minima of the standard deviations obtained from frequency-of-seeing curves are different for different *S*s. Once these data had been obtained, the question arose regarding their reliability. This then led to the development and the planning of Part 4 of the experiment.

TABLE 3

STANDARD DEVIATIONS OF THE LOGARITHM OF THRESHOLD ENERGY VALUES (*SD* LOG. $E_{.50}$) IN ARBITRARY UNITS FOR TWO *S*S AND 30 WAVELENGTHS (5mμ PASS BAND) OBTAINED FROM FREQUENCY-OF-SEEING CURVES

Wavelength (mμ)	*S* III	*S* IV
410	.139	.136
420	.113	.133
430	.122	.143
440	.122	.127
450	.100	.104
460	.118	.107
470	.132	.106
480	.109	.122
490	.111	.097
500	.121	.077
510	.126	.092
520	.106	.099
530	.118	.128
540	.142	.132
550	.113	.126
560	.151	.113
570	.135	.091
580	.128	.088
590	.128	.123
600	.130	.103
610	.130	.105
620	.122	.123
630	.180	.136
640	.125	.094
650	.123	.090
660	.149	.105
670	.120	.100
680	.134	.120
690	.205	.109
700	.140	.099

Part 4: Extensive Investigation of the Variability of the Standard Deviation of the Logarithm of Threshold Energy

In this phase of the investigation 36 wavelengths were studied. These were spaced at 10 mμ intervals from 410 to 700 mμ with the addition of 435, 445, 475, 485, 505, and 575 mμ. The stimulus light subtended 30 min. of arc. Two *S*s were employed. Two basic psychophysical methods were used; the method of constant stimuli and the method of limits. The method of constant stimuli was employed in the same fashion as in the preceding sections except that 50 blank stimuli were randomly interspersed with the 250 regular stimuli. The number of positive responses to blanks was used as an index of the attention and care with which the *S*s made their observations. The method of limits was employed in the following manner.

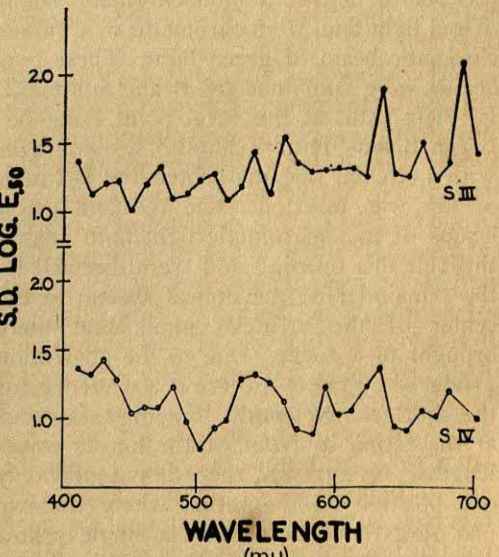


FIG. 5. Standard deviation (*SD* Log $E_{.50}$) values for two *S*s (method of constant stimuli).

Only the ascending phase of the method was used because it was desirable to avoid any possibility of changing the *S*'s state of dark adaptation. The method was used first on a straightforward threshold basis; then it was used on a basis of flicker response to a one-to-one light–dark ratio flash repeated at the rate of 20 flashes per second. For each type of threshold and for each wavelength, 25 determinations were made by each *S*. They were made, however, in five separate sessions, with five determinations at each of the wavelengths at each session. In order to control fatigue effects, the same rotational design employed by Dillon (1954) was used. In the first session, the *S* began with the 410 mμ wavelength and continued to the longer wavelengths until the 36 wavelengths had been tested. In the second session, the *S* began with wavelength 460 mμ and continued to 700 mμ, and then was tested from 410 to 450 mμ. In the third session, the *S* began at 520 mμ, and continued in the same fashion as in the preceding sessions. The fourth session began at wavelength 580 mμ, and the fifth at wavelength 640 mμ.

In order to secure greater accuracy of fixation, use was made of the phenomenon of chromatic parallax. Monochromator 4 was set to deliver a monochromatic beam of red light and Monochromator 3, a monochromatic beam of green light. These two beams were combined by Beam Splitter 2. In their path, at the focal point lying between Beam Splitters 2 and 3, a diaphragm with a circular opening subtending 1–2 min. of arc was inserted. The red and green beams of monochromatic light both passed through this opening and were directed by the remainder of the optical system to the center of the artificial pupil, and there brought to a focus. Due to the aberration produced by the eye, there is a tendency for the shorter wavelengths to come to a focus at the retina in front of the longer wavelengths. As a result, there is one and only one position on the retina where the two tiny spots coincide to give a single yellow dot. At all other positions two tiny spots appear—one red and the other green. In setting up the signal light, the *S* viewed the fixation point and adjusted his head and the chin rest until he was able to see a single yellow fixation point. The *S* then directed the experimenter who adjusted the position of the signal light field so that the fixation point fell in its center.

The results of this work, employing the method of constant stimuli, are contained in Tables 4 and 5, and in Fig. 6. An analysis of variance was undertaken with regard to the data obtained with the method of constant stimuli from both *S*s for 36 wavelengths and three sessions at each wavelength. The statistical model which fits this type of data is described by McNemar (1955) as Case VII, i.e., a mixed model wherein the rows stand for the different *S*s; the columns for the wavelengths; and there is measurement replication leading to three scores per cell. In this model the variance estimate of individual differences is used as the error term to test the significance of the interaction. The interaction variance is used to test the main effects of *S* and wavelength. The results of this analysis are given in Table 6. In this case, the interaction variance estimate as tested against the variance estimate of individual differences is not significant at the .05 level of confidence. Hence, the individual responses do not differ by more than can be expected on the basis of errors of measurement. Using the interaction estimate as the error term, it is found that the *S*s are significantly different one from the other. This is an expected result whenever measurements attain a reasonable degree of pre-

TABLE 4

Number of Affirmative Responses to "Blank" Stimuli

(Total Number of Curves—108 Per *S*)

S	Curves with Affirmative Responses			
	0	1	2	3
I	59	30	11	8
III	42	38	21	7

TABLE 5

STANDARD DEVIATIONS OF THE LOGARITHM OF THRESHOLD ENERGY VALUES (*SD* LOG $E_{.50}$) IN ARBITRARY UNITS FOR TWO *S*s AND 36 WAVELENGTHS (5 mμ PASS BAND) OBTAINED FROM FREQUENCY-OF-SEEING CURVES

Wave-lengths	*S* I				*S* III				Average Two *S*s
	Session			Average	Session			Average	
	1	2	3		1	2	3		
410	.194	.169	.154	.172	.166	.112	.122	.133	.1525
420	.182	.187	.183	.184	.181	.138	.114	.144	.164
430	.183	.187	.154	.175	.128	.145	.121	.131	.153
435	.174	.179	.155	.169	.157	.119	.143	.140	.1545
440	.182	.221	.179	.194	.137	.165	.113	.138	.166
445	.197	.179	.168	.181	.124	.131	.118	.124	.1525
450	.177	.123	.138	.146	.093	.145	.148	.129	.1375
460	.165	.164	.158	.162	.158	.114	.125	.132	.147
470	.156	.213	.143	.171	.110	.119	.114	.114	.1425
475	.127	.153	.118	.133	.146	.133	.118	.132	.1325
480	.148	.173	.160	.160	.171	.119	.130	.140	.150
485	.149	.132	.183	.155	.149	.134	.095	.126	.1405
490	.152	.190	.140	.161	.114	.125	.103	.114	.1375
500	.201	.114	.134	.150	.187	.115	.110	.137	.1435
505	.134	.142	.143	.140	.124	.128	.128	.125	.1335
510	.165	.125	.172	.154	.143	.132	.081	.119	.1365
520	.177	.122	.184	.161	.156	.117	.097	.123	.142
530	.186	.206	.179	.190	.119	.128	.119	.122	.1485
540	.161	.139	.119	.140	.107	.110	.099	.105	.1225
550	.187	.175	.115	.159	.082	.108	.086	.092	.1255
560	.149	.136	.154	.146	.167	.158	.098	.141	.1435
570	.199	.180	.171	.183	.160	.121	.130	.137	.160
575	.195	.145	.144	.161	.108	.120	.135	.121	.141
580	.156	.187	.145	.163	.133	.120	.166	.140	.1515
590	.211	.166	.146	.174	.131	.138	.146	.138	.156
600	.146	.192	.089	.142	.125	.147	.108	.127	.1345
610	.139	.148	.157	.148	.077	.103	.133	.104	.126
620	.155	.132	.155	.147	.110	.130	.139	.126	.1365
630	.155	.173	.136	.155	.079	.127	.140	.115	.135
640	.175	.217	.140	.177	.130	.155	.188	.158	.1675
650	.163	.136	.123	.141	.105	.134	.138	.126	.1335
660	.193	.121	.128	.147	.123	.142	.159	.142	.1445
670	.149	.212	.155	.172	.126	.095	.140	.120	.146
680	.145	.145	.166	.152	.129	.138	.134	.134	.143
690	.209	.186	.158	.184	.119	.101	.151	.124	.154
700	.136	.149	.127	.137	.116	.092	.120	.109	.123

cision. The *F* ratio of wavelength versus interaction yielded a value which just attained significance at the .05 level. In tracking down the causes of this significant ratio, two further analyses were carried out—one on each *S*. *S* III showed no significant differences. *S* I showed differences due to wavelength which were significant at the .01 level. The question of the more exact location of the differences was solved by means of Tukey's test for significance. Once wavelengths 420, 440, 445, 530, 570, and 690 mμ were removed from the matrix, the analysis failed to yield an *F* ratio significant at the .05 level. Study of the individual *SD* Log E scores for the above-mentioned wavelengths shows that there is in five of the six cases one large value

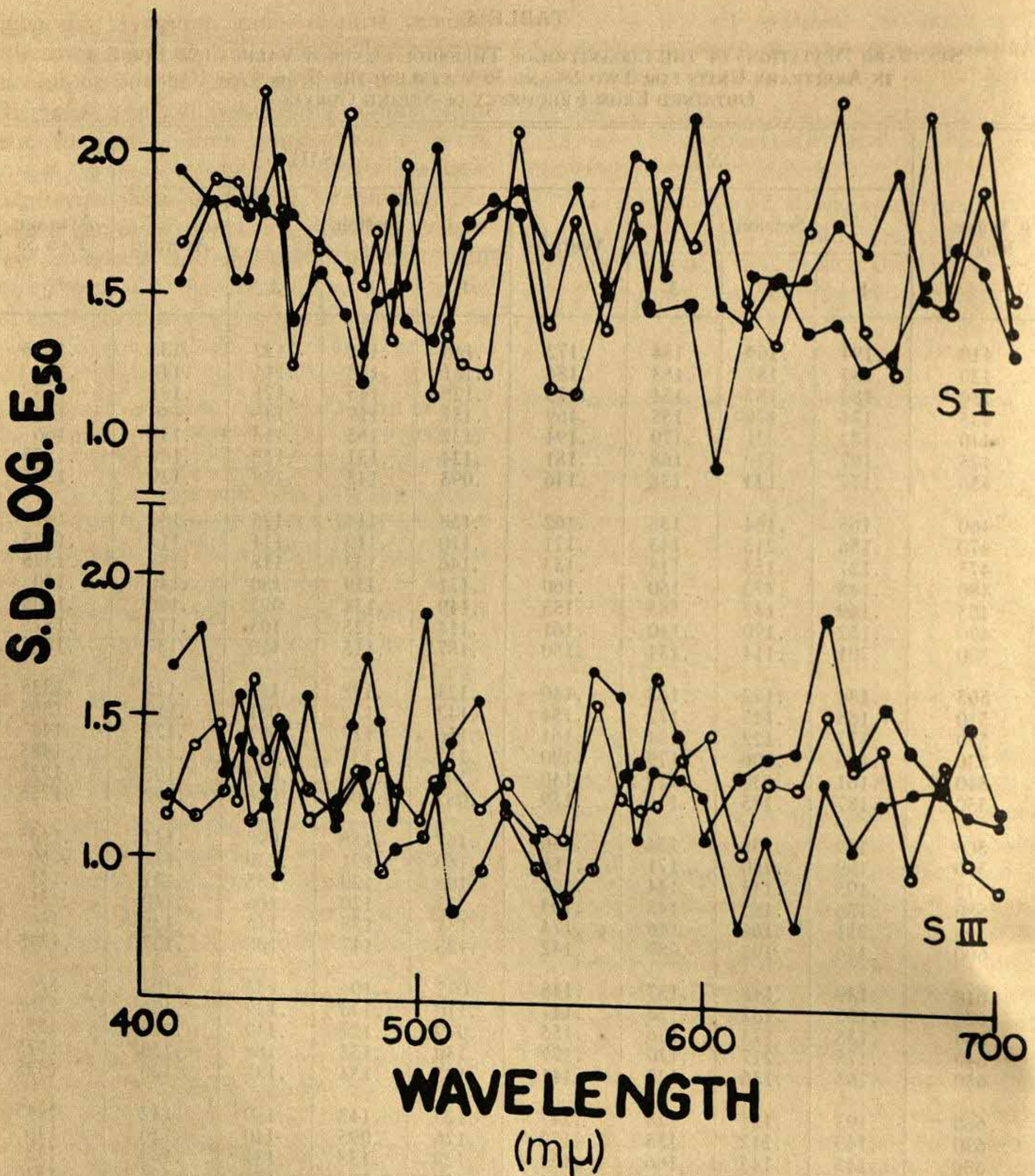


FIG. 6. Standard deviation (*SD* Log $E_{.50}$) values for two *S*s with three determinations per *S* per λ (method of constant stimuli).

together with two values of average magnitude. It is the large value which is producing the significant *F* ratio. There is nothing consistent in the values for the three sessions. In only one case, wavelength 420 m*μ*, is there consistency in large values for the three sessions. However, this trend is not borne out by *S* III. Intercorrelations between the scores for the three sets of data per *S* yield coefficients of .099, .097, .183 for one *S*, and .136, .067, .166 for the other. With this lack of consistency between the sets of measures it is concluded that no evidence is offered for a quasi-periodic variation of the *SD* Log E values with wavelength, as yielded by the method of constant stimuli.

After the work upon the magnitude of

TABLE 6

An Analysis of Variance of the Standard Deviation of the Logarithm of Threshold Energy Values According to the Variables of *S*s and Wavelengths with Three Replications

Source	*SS*	*df*	Variance Estimate	*F*
*S*s	60,200	4	60,200	127.00
Wavelength	29,279	35	836	1.76
Interaction	16,507	35	474	1.16
Individual difference	58,554	144	407	
Total	164,621	215		

the standard deviation, obtained by the method of constant stimuli, was completed, attention was given to determining the pattern of variation in the standard deviation when the ascending method of limits was used. As detailed in the procedure, each *S* gave five threshold determinations at each of 36 wavelengths during a session, and there were five such sessions, totaling 25 determinations per wavelength per *S*. Because interest centered on the magnitude of the standard deviation and any effect of day-to-day variability, contributed by the mean to this value, should be excluded, the standard deviation was obtained from the data in the following manner. The magnitude of the deviates was determined by subtracting the values of each five determinations from the average of the same five determinations. The 25 deviations thus obtained were squared and summed and divided by *N* and the square root extracted. It was felt that this would yield what might be called a refined measure of variability and one which would be uncontaminated by whatever causes produce the small but definite day-to-day variability in mean values reported in the literature. This would also give greater comparability to the standard deviations obtained by the method of constant stimuli where each standard deviation is determined as the variability around the mean value of the session from which it is taken. With this method of calculation, one finds that the variability is extremely small. The standard deviations obtained from the data, computed this way, are as small or smaller than those normally reported in the literature. In no case does a single standard deviation reach a value of .10 of a logarithmic unit. The data are given in Table 7 and Fig. 7.

In addition to using the strict threshold method, standard deviations were again obtained, using the ascending series of the

TABLE 7

Standard Deviations of the Logarithm of Threshold Energy Values (*SD* Log E) in Arbitrary Units for Two *S*s as Determined by the Ascending Method of Limits for Absolute Threshold and Critical Fusion Frequency Thresholds

Wavelength (mμ)	*S* I		*S* III	
	Threshold	C.F.F.	Threshold	C.F.F.
410	.064	—	.033	—
420	.049	—	.048	.030
430	.073	.045	.050	.026
435	.056	.051	.054	.043
440	.032	.052	.030	.048
445	.049	.036	.034	.053
450	.049	.052	.052	.036
460	.061	.062	.042	.052
470	.073	.041	.031	.052
475	.048	.063	.046	.040
480	.051	.052	.043	.046
485	.063	.064	.040	.047
490	.060	.034	.047	.046
500	.061	.056	.036	.050
505	.041	.040	.058	.051
510	.041	.038	.035	.040
520	.042	.050	.046	.046
530	.058	.042	.044	.030
540	.062	.062	.031	.047
550	.044	.071	.017	.044
560	.053	.048	.041	.033
570	.063	.059	.038	.055
575	.064	.043	.024	.051
580	.044	.067	.051	.056
590	.057	.045	.045	.046
600	.059	.051	.051	.052
610	.033	.073	.036	.050
620	.084	.049	.034	.039
630	.080	.055	.071	.085
640	.063	.084	.051	.049
650	.061	.054	.049	.089
660	.067	.041	.039	.082
670	.056	.047	.031	.041
680	.054	.052	.038	.044
690	.055	.050	.048	.029
700	.080	—	.049	—

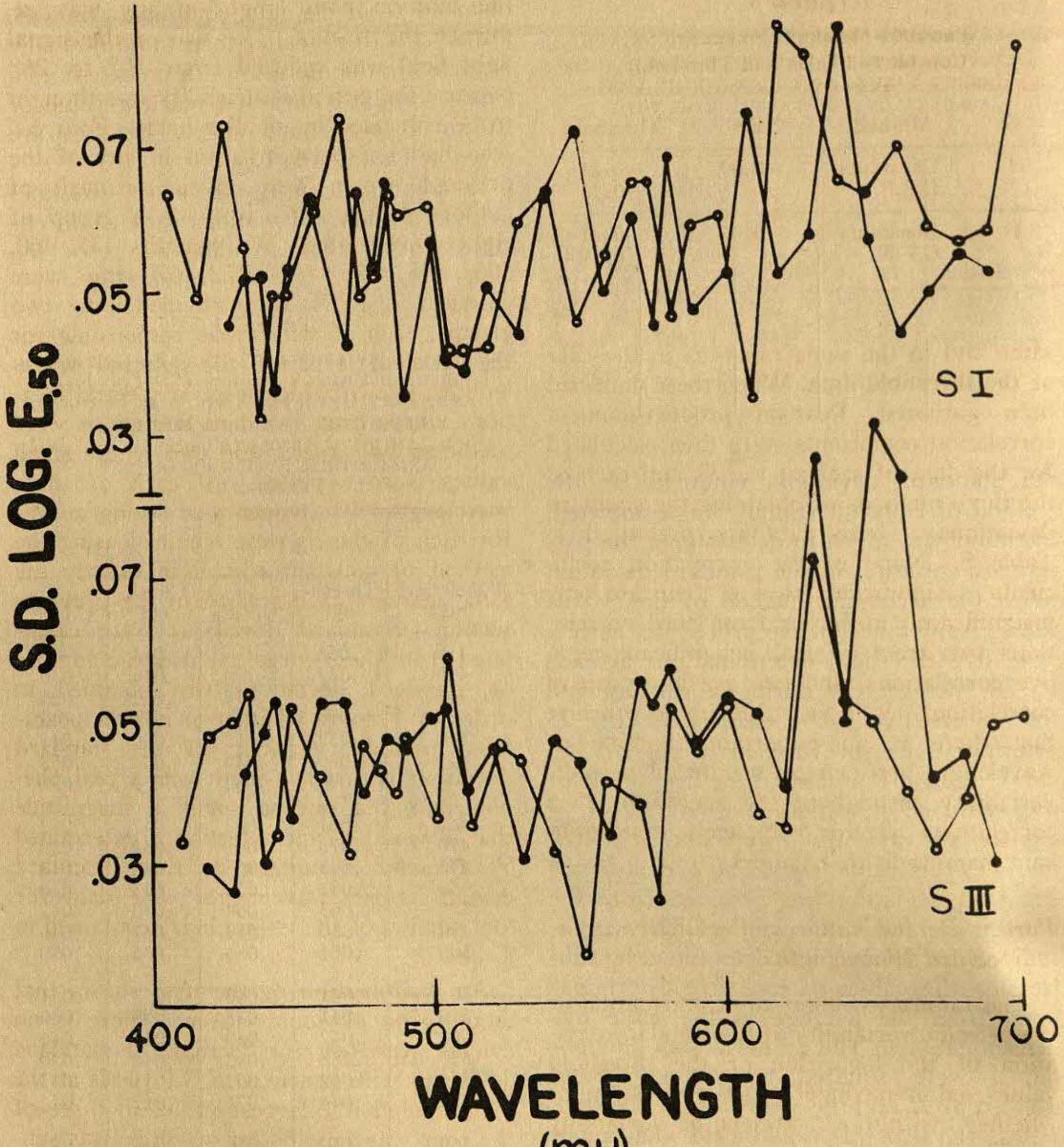


FIG. 7. Standard deviation (*SD* Log $E_{.50}$) values for two *S*s. Open circles, threshold; closed circles, C.F.F.

method of limits for obtaining the point of critical fusion frequency. This critical fusion frequency was obtained for a flash rate of 20 flashes of light per second, with a light–dark ratio of 1.0. This flash rate is sufficiently high to insure that the critical fusion frequency threshold will be at the cone level. The lack of determinations at wavelengths 410 mμ and 420 mμ for *S* I, and 410 mμ for *S* III, were due to the fact that the monochromator did not give sufficient light in the violet region of the spectrum for critical fusion frequency determination at this relatively high rate of flash frequency. These data are also given in Table 7. The standard deviations were calculated in exactly the same way as they were for the threshold data obtained by the method of limits. The magnitude of the standard deviations is approximately the

TABLE 8

Product-Moment Correlations
(Constants, Flicker and Threshold)

S	Method	C.F.F.	Threshold
I	Constants	−0.09	−0.21
	C.F.F.		−0.13
II	Constants	−0.03	0.31
	C.F.F.		0.19

same and in the same range as in the case of the threshold data. When these data had been gathered, Pearson product-moment correlation coefficients were then calculated for the data of each of the *S*s and each of the three methods of obtaining the standard deviations. These data are presented in Table 8. None of the correlation coefficients is significant; two of them are only insignificantly different from zero correlations, two are very small insignificant negative correlations, and two are low positive correlations. It is very difficult to conceive that there is an important function of wavelength producing a significant periodic variability throughout the spectrum when correlations of this completely nonsignificant magnitude are obtained.

Part 5: Determination of Variability at Specified Wavelengths

The failure to obtain any confirmation of the periodic variability of the standard deviation of the logarithm of the threshold values led immediately to the question of whether, within a wavelength, variability of .20 of a logarithmic unit was excessive and must be attributed to some factor other than session-to-session variability. Hence, Part 5 of this research was designed.

The apparatus and general procedure was the same as for the constant stimuli section of Part 4 with the exception of two minor changes in the apparatus. The fixation point was changed from a tiny dot in the center of the signal light to one which consisted of two very small dots placed one on each side of the signal light. This was done in order to remove the possibility that the fixation point might in any way influence the results. The size of the signal light field was reduced from .50° to .26° because the periodic variability was thought to be enhanced by smaller fields. Four *S*s, who had not been involved in any of the previous work, were given the task of gathering data extensively on a group of eight wavelengths. Wavelengths 435, 460, 506, 545, 575, 609, 622, 670 $m\mu$ were chosen. The *S*s were divided into two groups, each of which was responsible for the data of four of the selected wavelengths, two of which were in spectral locations where large standard deviations were predicted by Crozier and two where small values were expected. At each of these wavelengths 10 frequency-of-seeing curves for each of the *S*s were obtained using the method of constant stimuli in exactly the same manner as described in the previous section. Standard deviations were calculated from these curves. This large number of standard deviations was obtained in order to discover whether or not the postulated periodic variation of the standard deviation with wavelength was a real phenomenon, but one so small in magnitude that it would escape detection if determined by as small a number as three standard deviations per wavelength. The data for this portion of the research are contained in Table 9.

An examination of the data shows that there is no obvious pattern. There is no evident tendency for decrease of standard deviation with practice. *S* VI yields at the fifth session, the largest values in three of the four wavelengths for which he was subject. *S* VII yields his largest values at the last session, second session, first session, and the sixth session. *S* VIII has his largest values once in the last session, and the other three in the first two sessions. It is also to be noted that the values of the standard deviations are of approximately the same magnitude as those obtained by *S*s I, II, III, and IV. The size of the standard deviation ranges from .085 to .263 log units, with less than 10% of the total number attaining values of .20 or larger. It seems difficult to draw any conclusions

TABLE 9

Standard Deviation of the Logarithm of Threshold Energy Values for Eight Wavelengths, Four *S*s, and 10 Sessions per Wavelength

Session	Wavelengths							
	435 mμ	460 mμ	506 mμ	545 mμ	575 mμ	609 mμ	622 mμ	670 mμ
	S V				*S* VII			
1	.201	.137	.174	.135	.125	.120	.255	.112
2	.162	.163	.181	.118	.104	.221	.155	.130
3	.173	.182	.135	.156	.170	.125	.150	.133
4	.234	.215	.189	.138	.146	.152	.164	.116
5	.120	.214	.177	.149	.119	.168	.161	.107
6	.165	.167	.137	.133	.133	.150	.117	.157
7	.234	.183	.118	.150	.124	.131	.165	.090
8	.204	.249	.224	.128	.115	.125	.121	.112
9	.240	.217	.181	.162	.137	.123	.111	.110
10	.263	.183	.181	.1781	.177	.119	.158	.100
Mean *SD*	.200	.191	.170	.145	.135	.143	.156	.117
Range	.143	.112	.089	.060	.073	.102	.144	.067
Session	*S* VI				*S* VIII			
1	.168	.128	.099	.114	.163	.221	.161	.122
2	.132	.109	.127	.139	.135	.220	.152	.164
3	.110	.085	.109	.117	.096	.170	.097	.141
4	.132	.116	.156	.106	.135	.126	.142	.170
5	.164	.223	.180	.156	.129	.146	.137	.144
6	.165	.131	.123	.114	.160	.115	.111	.103
7	.160	.148	.169	.087	.127	.142	.087	.122
8	.137	.161	.104	.113	.159	.132	.124	.139
9	.136	.160	.123	.096	.132	.110	.133	.107
10	.159	.197	.135	.112	.169	.160	.141	.137
Mean *SD*	.146	.146	.133	.115	.141	.154	.129	.135
Range	.055	.138	.081	.069	.073	.111	.055	.067

from these data beyond the fact that the variability from session to session is of approximately the same order for all four *S*s. It is felt that this phase of the experiment represents a fairly good sample of *S* variability in the task of obtaining frequency-of-seeing curves under constant conditions of wavelength, field size, field position, and time of exposure.

An inspection of the data shows that the range of 10 standard deviations within one wavelength and for one *S* is less than .150 log units at the maximum and .055 at the minimum, with other values assuming intermediate positions. The important conclusion to be drawn from these data is that a group of standard deviations obtained during a series of sessions yields values whose ranges amount to between 25% and 50% of the values assigned by Crozier (1950) as representing a real shift in the standard deviation as a function of wavelength. The pattern of the average standard deviations (averaging all standard deviations for both *S*s and for each wavelength) shows a trend of a steady decrease from the violet end of the spectrum to the red. It is suggested that this decrease should be interpreted in line with the generally established fact that variability from session to session and from *S* to *S* is greater in the shorter wavelengths than in the longer. It would seem, then, that the standard deviation reflects directly the magnitude of the

TABLE 10

AVERAGE VALUES (FROM 15 FREQUENCY-OF-SEEING CURVES) OF THE STANDARD DEVIATION OF THE ENERGY AT FOUR WAVELENGTHS AND THREE FIELD SIZES, AND FOR TWO *S*s[a]

Field Size	*S*	Wavelengths			
		460 mμ	506 mμ	546 mμ	575 mμ
3 min.	I	.172 (.0238)	.168 (.0175)	.128 (.0208)	.143 (.0225)
	II	.216 (.0340)	.168 (.0169)	.161 (.0209)	.161 (.0175)
30 min.	I	.158 (.0137)	.165 (.0193)	.155 (.0211)	.156 (.0168)
	II	.170 (.0223)	.178 (.0179)	.146 (.0249)	.163 (.0168)
45 min.	I	.160 (.0250)	.136 (.0150)	.141 (.0171)	.159 (.0132)
	II	.150 (.0194)	.121 (.0154)	.126 (.0166)	.130 (.0208)

[a] Entries in parentheses represent the *SD*s of the *SD*s taken as raw scores (arbitrary Log E units).

variability of the mean values. If this interpretation is correct, Crozier's (1950) contention that visual threshold values are best described in terms of a logarithmic Gaussian distribution is not borne out. One of the reasons presented by him, as decisively proving that a Poisson distribution could not describe visual threshold data, was that the standard deviation varied periodically with wavelength, and that its magnitude was completely independent of the magnitude of the mean value. The data obtained from this research in no way support this contention. Hence, whether a Poisson distribution, an arithmetical Gaussian distribution, or a logarithmic Gaussian distribution represents the best mathematical model for the purpose of describing threshold data obtained by the method of constant stimuli, is still a problem requiring determination. This same conclusion has been reached by Blackwell (1953).

Part 6: Further Determination of the Variability of the Standard Deviation With Changing Field Size

The last aspect of the study of the variation of the standard deviation involved its magnitude as a function of field size. Four 5-mμ bands of monochromatic light were again used. These bands were centered on 460, 506, 546, and 575 mμ and were chosen because two of them represented spectral bands where maximal values and two where minimal values of the standard deviation were to be found according to the data of Crozier (1950). Because there was evidence that field size might have an effect on the function, three field sizes were used which subtended, at the cornea, visual angles of 3, 30, and 45 min. of arc. The method of constant stimuli was used in exactly the same manner as in Parts 4 and 5 to obtain frequency-of-seeing curves. Two *S*s each obtained 15 of these curves for each combination of wavelength and field size.

The mean and standard deviation for each frequency-of-seeing curve was calculated by means of the Muller-Urban system of weights. Table 10 gives the average values of the standard deviations obtained in this experiment. Each entry represents the average of 15 standard deviations. The measure is the standard deviation of the logarithm of the energy expressed in arbitrary units.

A three-way analysis of variance employing these average values as entries was calculated. The main effects studied were wavelength, field size, and *S*s. Table 11 gives the pertinent data yielded by the analysis. This experimental plan comes under McNemar's classification of a mixed model, and is similar to his Case XIV (McNemar, 1955). Following this guide, the triple interaction value was used to test

TABLE 11

ANALYSIS OF VARIANCE FOR THE AVERAGE STANDARD DEVIATION VALUES FROM FOUR WAVELENGTHS, THREE FIELD SIZES, AND TWO *S*s

Source	*SS*	*df*	Variance Estimate
Wavelength	2,518	3	839
Field Size	2,847	2	1,424
*S*s	81	1	81
Wavelength × Field Size	1,598	6	266
Wavelength × *S*s	291	3	97
Field Size × *S*s	1,650	2	825
Wavelength × Field Size × *S*s	531	6	89

the two-way interactions; the wavelength-*S* interaction was used to test the effect of wavelength, and the field size-*S* interaction was used to test the effects of field size. The *F* ratios show significance at the .05 level for field size-*S* interaction. All other interactions are insignificant at this level of confidence. Among the main effects, neither field size nor wavelength effects are significant at the .05 level. This conclusion is in line with the findings of previous research in this laboratory, but directly at variance from the data of Crozier (1950). The data reported here are those of two *S*s. In previous sections the work relative to this same matter was described. This involved six *S*s over and above the present two. All the results are in agreement. Since Crozier's data were obtained from a single *S*, it would seem that the weight of evidence at the moment is against the proposition that the magnitude of the standard deviation is influenced by wavelength.

TABLE 12

ESTIMATED NUMBER OF QUANTA FOR THRESHOLD (50% POINT ON FREQUENCY-OF-SEEING CURVES) FOR FOUR WAVELENGTHS, THREE FIELD SIZES, AND TWO *S*s

Field Size	*S*	Wavelength			
		460 $m\mu$	506 $m\mu$	546 $m\mu$	575 $m\mu$
3 min.	I	8,484	1,945	994	1,306
	II	12,078	2,598	1,266	1,261
30 min.	I	75,935	23,449	9,807	12,741
	II	84,396	25,882	13,235	17,221
45 min.	I	154,260	32,290	15,851	15,846
	II	180,240	54,324	35,845	35,844

Part 7: Quantal Estimates of Mean Values Obtained at Four Wavelengths

The frequency-of-seeing curves described in the preceding section supplied a large amount of data the mean values of which had been calculated in radiometric units, namely, microwatts. It was entirely a matter of calculation to change these units into quanta. The formula used was:

$$\frac{\text{Energy in Ergs}}{h\nu}$$

where h is Planck's constant and ν is the frequency of the light. The values were also corrected for the duration of the flash which was .02 sec. The estimates are, therefore, of the total number of quanta delivered in the flash. They are given in Table 12 and represent that number of quanta at the cornea at which the stimulus is seen 50% of the time. With these data as a base, further calculations were carried out in order to discover the number of quanta which actually entered into the photoreceptors and produced the threshold sensation.

Because aberration is a necessary concomitant of light flashes delivered through optical systems with small apertures, correction for this was the first step in an evaluation of the quanta-receptor ratio. Five different approaches to the problem offered themselves. The first one was derived from a statement made by Nolan (1957) that Karn's (1936) data on the relationship of time and intensity show that aberration increased the size of his fields by 7.3–7.4 min. of arc. Though the fields in this experiment were smaller than the two larger ones employed here, calculations were carried out using a correction factor of 7 min. of arc. The retinal field diameters were then assumed to subtend angles of 10, 37, 52 min. of arc. Using these retinal areas together with the cone density ratios calculated from Polyak's (1941) data, the ratios of quanta per cone were derived.

In addition to the calculations carried out according to the theories of Nolan (1957), two further analyses were done, each according to two slightly differing assumptions. If a piece of frosted glass is placed at the artificial pupil of the apparatus, the light pattern appearing in the pupil has the form of a vertical bar of light across the center of the artificial pupil. The height of the light bar is determined by the diameter of the artificial pupil which is 2.5 mm. The width of the bar is a function of (*a*) the width of the exit slit of the monochromator, and (*b*) the magnification of the optical system. This latter is determined by the focal length of the lenses in the beam of light. The collimating lens in the front of the monochromator has a focal length of

11 in. The converging lens and the negative lens, located in proximity to the artificial pupil, yield, in combination, a focal length of 15 in. Hence, the magnification factor of the system is 1.36364.

The first formula to be applied to the data is the same as that used by Wright (1947).

$$\theta = \frac{2.44 \times \lambda \times 10^{-3}}{NP'}$$

where θ is the angular subtense to the first ring of the Airy diffraction disc, λ is the wavelength in microns, N is the refractive index of the vitreous humor—equal to 1.34, P' is the width or the length, respectively, of the light patch at the artificial pupil. Width equals slit width × 1.34354. Length equals the diameter of the artificial pupil, i.e., 2.5 mm.

A second method of determining the size of the angle is the use of the ratio

$$\text{sine } \theta = \frac{\lambda}{w}$$

where λ is the wavelength in microns and w is the width or length of the light patch at the artificial pupil.

Judging from the fact that what the *S* sees is reported as a circle of light, the diffraction pattern on the retina must be some sort of an ellipse whose distortion is so small that it is not appreciated by the *S* as being really different from a circle. Certainly the greatest possible amount of distortion would be that of a rectangular patch of light. In order to cover all possibilities, both of the formulae just mentioned were used to calculate the height and the width of the patch of light. Then these two dimensions were used to complete the area of a rectangle and an ellipse according to the standard formulae. Again Polyak's (1941) data was employed to derive quanta-per-cone ratios.

The next factor to be considered in this process of determining the effective number of quanta at the receptor level is that of the absorption of light both by the ocular media and, since these data are foveal, by the macular pigment. The quantum-per-cone ratios were corrected using the transmission factors of Ludwig and McCarthy and of Wald as tabulated by Judd (1951).

The final correction which had to be made was for the effective absorption of the quanta by the photochemical substance within the cones. It is not known exactly what this factor is, but it is commonly held that it is less than the absorption factor of visual purple, which is probably in the neighborhood of 10% (Pirenne & Denton, 1952).

The results of these calculations as applied to each method of correction for diffraction are contained in Table 13. The two columns headed No. 1 contain the quanta-per-cone ratios without any corrections made for diffraction. Columns No. 2 and No. 3 give the quanta-per-cone ratios using the formula from Wright; columns No. 2, rectangle; columns No. 3, ellipse. Columns No. 4 and No. 5 give the results when employing the sine formula with columns No. 4, rectangle, and columns No. 5, ellipse. Columns No. 6 give the results using Nolan's (1957) suggested correction. These data show ratios of from 51 to 6 quanta per 100 cones. The higher ratios are concentrated especially at wavelengths of 460 and 506 mμ and the 3-min. field. This pattern is consistent with the decreased sensitivity at the foveal center for the shorter wavelengths in the visual spectrum. The differences between the ratios for these two wavelengths with the 3-min. field and the 45-min. field are of the same order found by Sperling and Hsia (1957) and by Wright (1947).

Polyak (1941) speaks of two types of cone systems in the foveal area. Anatomically these systems are for the most part identical in their component structures. However, they are undoubtedly different in their functioning. They are called, respectively, the common, or multisynaptic, system and the private, or monosynaptic, system. The midget bipolar (h) and the midget ganglion cell (s) are essential to the functioning of the monosynaptic system. The common system can function without their involvement. Since it is well known that visual acuity is proportional to the intensity of the stimulating light, most probably the monosynaptic system is the mecha-

TABLE 13

Quanta Per Cone Ratio Without Corrections for Aberration and With Correction for Same (Correction for Photo Pigment Absorption Included)

Field Size	Wave-length	*S* I							*S* II						
		1	2	3	4	5	6	Av.	1	2	3	4	5	6	Av.
3 min.	460	1.4	.27	.35	.28	.36	.10	.27	2.0	.39	.50	.40	.51	.14	.39
	506	1.2	.20	.25	.16	.20	.09	.18	1.6	.26	.33	.21	.27	.14	.24
	546	1.0	.12	.16	.10	.12	.08	.12	1.5	.16	.21	.12	.16	.09	.15
	575	1.3	.14	.18	.12	.16	.10	.14	1.3	.14	.17	.12	.15	.10	.14
30 min.	460	.2	.12	.15	.11	.14	.11	.13	.2	.13	.17	.12	.15	.12	.14
	506	.2	.14	.18	.12	.15	.13	.14	.2	.15	.18	.12	.15	.13	.15
	546	.1	.09	.11	.07	.09	.08	.09	.2	.12	.15	.10	.12	.12	.12
	575	.2	.10	.13	.10	.13	.12	.12	.2	.14	.18	.13	.17	.16	.16
45 min.	460	.2	.12	.15	.10	.13	.11	.12	.2	.14	.17	.12	.15	.13	.14
	506	.2	.09	.11	.08	.10	.09	.09	.3	.15	.19	.13	.17	.15	.16
	546	.1	.07	.09	.06	.07	.07	.07	.3	.15	.19	.13	.17	.16	.16
	575	.1	.06	.08	.06	.08	.08	.07	.3	.15	.19	.14	.18	.17	.15

Note.—See text for explanation of numbers at head of columns.

nism whereby high visual acuity is mediated. It would not be stretching an hypothesis too far to assume that the (h) bipolar and the (s) ganglion cell mediate impulses at light intensities corresponding to a high index of visual acuity. Since the mop bipolar (d) is apparently not involved in foveal or cone vision, only the (e) and (f) [(g) is probably a freak form of either (e) or (f)] are functioning in foveal threshold vision.

Excluding the short wavelengths and the smallest field, a flux density of one quantum per 6–14 cones is required for threshold responses at the central portion of the luminosity curve. The (e) and (f) bipolars are related to the cones in a ratio of one bipolar for every 6–12 cones. This brings up the interesting hypothesis that the flux density in quanta is related to the diffuse types of foveal bipolars in approximately a one-to-one ratio.

Summary

The results of this investigation show:

1. There is no evidence for the process of photosensitization insofar as the four chosen wavelengths are concerned.

2. There is no evidence for the variation, as a function of wavelength, in the magnitude of the standard deviation of the logarithm of threshold energy as indicated by: (*a*) an investigation at 36 wavelengths by three psychophysical methods, (*b*) an extensive investigation of eight wavelengths by the method of constant stimuli, (*c*) an extensive investigation of four wavelengths with three different field sizes.

3. The magnitude of △I in relation to the background is of exactly the same value as the absolute threshold for that particular wavelength, until the background has attained an intensity which is higher than the cone threshold.

4. In conformity with the statements of Denton and Pirenne (1954), it must be assumed that day-to-day variability over a long period of time is of a magnitude of approximately .20 of a logarithmic unit.

5. The estimated number of quanta required for foveal stimulation after appropriate corrections have been made lies between 10–50 quanta per hundred cones at the 3-min. field, between 18 to 7 per hundred cones at the 30-min. field, and between 19 to 6 per hundred cones at the 45-min. field.

REFERENCES

BLACKWELL, H. R. Studies of the form of visual threshold data. *J. Opt. Soc. Amer.*, 1953, **43**, 456-463.

CROZIER, W. J. On the visibility of radiation at the human fovea. *J. gen. Physiol.*, 1950, **34**, 87-136.

DENTON, E. J., & PIRENNE, M. H. The absolute sensitivity and functional stability of the human eye. *J. Physiol.*, 1954, **123**, 417-441.

DILLON, D. J. An investigation of threshold variability in the determination of the photopic luminosity curve. Unpublished manuscript, Fordham Univer. Library, 1954.

HECHT, S. Vision II: The nature of the photoreceptor process. In Murcheson (Ed.), *A handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934. Pp. 704-828.

HECHT, S. Energy and vision. *Amer. Scientist,* 1944, **32**, 159-177.

HSIA, Y., & GRAHAM, C. H. Spectral sensitivity of the cones in the dark adapted human eye. *Proc. Nat. Acad. Sci.*, 1952, **38**(1), 80-85.

HURVICH, L. M., & JAMESON, DOROTHEA. Spectral sensitivity of the fovea: I. Neutral adaptation. *J. Opt. Soc. Amer.*, 1953, **43**, 485-494.

JUDD, D. B. Basic correlates of the visual stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 811-867.

KARN, H. W. Area and the intensity-time relation in the fovea. *J. gen. Psychol.*, 1936, **14**, 360-369.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

NOLAN, G. F. On the functional relation between luminous energy, target size, and duration for foveal stimuli. *J. Opt. Soc. Amer.*, 1957, **47**, 394-397.

PIRENNE, M. H., & DENTON, E. J. Accuracy and sensitivity of the human eye. *Nature,* 1952, **170**, 1039.

POLYAK, S. L. *The retina.* Chicago: Univer. of Chicago Press, 1941.

SPERLING, H. G., & HSIA, H. Some comparisons among spectral sensitivity data obtained in different retinal locations with two sizes of foveal stimulus. *J. Opt. Soc. Amer.*, 1957, **47**, 707-713.

WALD, G. Photochemistry of vision. In F. P. Fischer, A. J. Schaeffer, & A. Sorsby (Eds), *Documenta opthalmolgica,* Vol. **3**. 'S-Gravehage, Hollandia: Dr. W. Junk, 1948. Pp. 94-137.

WRIGHT, D. W. *Researches on normal and defective color vision.* St. Louis: Mosby, 1947.

(Accepted for publication September 16, 1958)

Vol. 73, No. 12 Whole No. 482, 1959

Psychological Monographs: General and Applied

CORRELATES OF LATER PERFORMANCE AND SPECIALIZATION IN PSYCHOLOGY

A FOLLOW-UP STUDY OF THE TRAINEES ASSESSED IN THE VA SELECTION RESEARCH PROJECT[1]

E. LOWELL KELLY
University of Michigan
AND
LEWIS R. GOLDBERG
Stanford University

BACKGROUND

As World War II drew to a close, the Veterans Administration—faced with the immense challenge of rehabilitating a returning army of men, many of whom would need psychiatric services—recognized the extreme shortage of professional personnel trained to perform these services and subsequently instituted training programs in pertinent professional fields to help increase the supply of such specialists. In 1946, the Veterans Administration began its training program in clinical psychology and simultaneously sponsored a program of research aimed at increasing the effectiveness of the procedures used to select and train clinical psychologists.

This extensive research program (called the VA Selection Research Project), was directed and coordinated through the Psychology Department of the University of Michigan and carried out on a nation-wide basis between 1946 and 1951. The first year was essentially devoted to planning and carrying out three preliminary assessment programs (in Ann Arbor, Michigan; Farmingdale, New York; and Asilomar, California). During these preliminary assessments, VA trainees were used as *S*s in an attempt to develop and refine techniques which were to be used in the later major assessments.

The summer of 1947 marked the first of these major assessment programs, this in Ann Arbor, Michigan. One hundred and thirty-seven incoming graduate students who had just been admitted to the VA training program in clinical psychology came to Ann Arbor, approximately 24 at a time, from all over the United States. Each group was subjected to an intensive series of assessment techniques during a five-day stay at the assessment center.

The assessment procedures included a battery of objective tests and another battery of diagnostic procedures that were more clinical in nature. Such objective instruments as the Strong Vocational Interest Blank, the Miller Analogies Test, the Minnesota Multiphasic Personality Inventory, the Guilford-Martin Battery of Personality Inventories, the Allport-Vernon Scale of Values, and the Thurstone Test of Primary Mental Abilities were administered to all of these *S*s.

---

[1] Early publication of this monograph was made possible by a special allocation to the Ann Arbor VA Hospital from the Education and Research Budget of the VA Central Office. The authors wish to acknowledge the encouragement and assistance received from Philip A. Smith, formerly of the Ann Arbor VA Hospital, as well as the help of the rest of the hospital's psychological staff. Special thanks are due to Antonio Monteiro for the many months of work which he devoted to the analysis of the data. The salaries of Mr. Monteiro and the junior author were paid in part by the Veterans Administration. Statistical analysis was carried out with the help of an IBM 650 Computer, which was freely made available by the Statistical Laboratory of the University of Michigan. Office supplies and clerical help were provided by the Psychology Department of the Ann Arbor VA Hospital and the Bureau of Psychological Services of the University of Michigan. And finally, we wish to thank our 245 *S*s, whose contribution of their time and thought made this study possible.

The less-structured procedures were used primarily to provide assessment staff members with bases for making descriptive, evaluative, and predictive judgments about the candidates. The techniques included two varieties of interviews, a series of situational tests patterned after the O.S.S. assessment model (O.S.S. Assessment Staff, 1948), and four projective tests (the Rorschach Inkblot Technique, the Thematic Apperception Test, the Bender-Gestalt Test, and a sentence completion test). In addition, each trainee wrote a detailed autobiography and as a supplement filled out a more-structured Biographical Inventory. Besides these sources of information, each staff judge had access to a credential file consisting of the trainee's application forms to the VA program and to the universities of his choice, any correspondence with these universities, his letters of recommendation, and his college transcripts.

Thirty experienced clinicians—mostly clinical psychologists, but with a few psychiatrists and social workers—served as staff judges during this assessment, and each made qualitative and quantitative evaluations on the basis of the particular assessment techniques with which he had contact. Great care was taken to insure that evaluations based upon one technique (or group of techniques) would not be "contaminated" by information secured from other sources (i.e., some other technique, an informal contact with the trainee, etc.). Finally, after all the individual evaluations were completed, all judges who had been concerned with a trainee came together and, in conference, pooled their judgments to form an additional composite qualitative and quantitative picture of each *S*.

Because all the *S*s utilized in this assessment had been notified of their admission to graduate schools and to the VA training program, the need was felt for an additional assessment to be secured under *applicant* conditions. Accordingly, in 1948 another nation-wide assessment program was instituted, this one using as *S*s all applicants to the clinical psychology programs of accredited universities throughout the country. These applicants were tested in special centers set up for this purpose all over the United States, and the protocols of their tests were then forwarded to a central headquarters. After all applicants were tested and graduate schools had made their selections, these data were evaluated by clinical psychologists in a manner similar to that employed in the 1947 assessment.

There were two essential differences between the two major assessment programs: (*a*) *S*s in the 1947 program were evaluated on the basis of techniques administered *after* admission to graduate school, while in 1948 all testing was carried out *prior* to such acceptance, and (*b*) in 1947 staff judges had face-to-face contact with the *S*s while the 1948 judges worked without benefit of this interaction. For the latter reason, such techniques as interviews and situational tests were not available to the 1948 judges.

On the other hand, many techniques were common to the two assessment programs. Applicants in 1948 took such objective tests as the Miller Analogies Test, the Strong Vocational Interest Blank, the Guilford-Martin Battery of Personality Inventories, and the Allport-Vernon Scale of Values in addition to such projective instruments as the Rorschach Ink-blot Technique and a sentence completion test. Again, staff judges had access to the complete credentials file for each applicant as well as to an autobiography and a Biographical Inventory. On the basis of these materials, evaluative ratings were made for 98 persons who were accepted by university departments and the VA program and for 21 others who, out of the 545 who were completely tested upon application, were refused admittance to this program.

The second half of the project was devoted to (*a*) the development of a battery of criterion measures, (*b*) the administration of these measures to trainees near the end of the four-year training program, and (*c*) the analyses of the interrelationships among the predictor and criterion measures. The findings were reported in Kelly and Fiske (1951). Since the Kelly-Fiske book is now out of print, and in order to provide present readers with a review of the findings of the original study, a section of the summary chapter is quoted in Appendix A.

The 1957 Follow-up Study

In their summary chapter, Kelly and Fiske said:

> In evaluating the findings of this project, it must be remembered that all of the criterion measures were collected relatively early in the professional lives of the subjects studied, and that the findings here reported are based on these early criteria of professional success. It is, of course, entirely possible that predictors which look unpromising at this time may have greater validity for predicting later criteria of professional accomplishment and *vice versa.* Although the writers have no immediate plans for further follow-up studies, they hope to be able to obtain even more meaningful criterion measures of the subjects after 10, 15, or 20 years and to analyze the validities of all predictors against criteria of mature professional success (p. 197).

A number of considerations led to the decision not to wait even 10 years before undertaking a follow-up study to obtain later and very different indices of professional performance: The original study had gradually been assimilated into the mainstream of assessment lore, and decisions regarding the selection and training of graduate students could be influenced by implications from this study. For this reason, it seemed desirable to have an early follow-up check on the later validity of some of the earlier findings, especially where definitive data were just not available in 1951 (i.e., eventual academic attainment) and where conclusions could have been reached prematurely. Secondly, the rapid and marked changes in the nature of clinical psychology as a specialty leads to interest in the present satisfaction of *S*s who originally entered the field when conditions were somewhat at variance from the present state of the profession. In addition, the relative shortage of professional personnel over the years and the consequent job mobility of the original *S*s furnishes a rich background for a study of specialization within the field of clinical psychology. And finally, our colleagues and many of the *S*s repeatedly urged us to study post-PhD criteria.

Thus, early in 1957, we decided to conduct the first follow-up study of the trainees originally assessed in 1947 and 1948. Having made this basic decision we gave considerable thought to its nature, scope, and methodology. For reasons sentimental as well as scientific, we considered inviting all of the original *S*s to return to Ann Arbor for another large-scale assessment program or of sending trained crews of assessors all over the United States to obtain performance criteria of former trainees in their present work settings. Appealing as each was, either program would have been extremely expensive. But, even had unlimited funds been available, it is not likely that either of them would have been feasible. Most of the *S*s now hold responsible positions and are hence not able to contribute as much time as would have been necessary. Furthermore, because of the degree of specialization within the group, it is probable that many would have been resistant to the prospect of being evaluated with respect to skills no longer practiced. Finally, since the intervening years did not yield measures of diagnostic and therapeutic skills superior to the ones employed in the original study, it simply did not seem advisable to try to evaluate professional competence again at this time. For these reasons, it was decided that this first follow-up would be a modest one, utilizing new criteria obtainable from readily available sources.

THE FOLLOW-UP QUESTIONNAIRE

One available source, the American Psychological Association *Directory,* contained some information for those original *S*s who had become APA members. Certain additional information was obtainable from the departments of psychology in the universities in which they had begun their graduate training. Most promising, however, was the potential utility of a questionnaire. The original *S*s were a highly selected group, capable of intelligent cooperation in a mail follow-up and probably sufficiently interested to provide much of the same information that an interview would reveal. There seemed to be little value in asking for self-evaluations of present professional competence, but we felt certain that the *S*s could, and would, tell us about their professional work history, their work preferences as of today, their preferences for future work, and also report present satisfaction with their earlier professional choices. For several obvious reasons, both ethical and technical, we promised not to secure evaluations from their present employers or supervisors.

Administration of the Follow-up Questionnaire

In 1957, we began a follow-up study of the 245 men on whom assessment evaluations were completed in 1947 and 1948. Information concerning the academic achievements of these *S*s was obtained from the universities in which they had been enrolled, and each *S* was requested to fill out a brief questionnaire pertaining to his professional history since leaving graduate school. A copy of this questionnaire—designed to provide information concerning the *S*s' academic and occupational histories, their work preferences and intentions, and their satisfaction with their professional choices—can be found in Appendix B.

Approximately 80% of the original *S*s were members of the American Psychological Association in 1957, and their addresses were available from this source. Attempts were made to locate the remaining 20% of the *S*s through old addresses, college alumni associations, telephone directories from the major cities in the United States, etc. Letters and questionnaires were repeatedly sent to each of the *S*s for whom addresses were available until they returned a completed questionnaire.

TABLE 1

ACADEMIC PROGRESS OF MALE ASSESSEES BY 1957

1947 Assessment: Number assessed = 128
1948 Assessment: Number assessed = 120
Total assessed = 248

Of these 248 persons:
- 3 died during the interval 1947–1957.
- 5 decided not to enter graduate school in psychology.[a]
- 21 were not admitted to the VA program in 1948 ("Rejects").

Thus leaving 219 living ex-trainees. Of these 219 persons:
- 154 (70%) achieved a PhD degree in psychology.
- 32 (15%) achieved an MA degree in psychology.
- 12 (6%) are still graduate students in psychology.[b]
- 5 (2%) achieved an MD degree.
- 5 (2%) achieved PhD or MA degrees in other fields.
- 11 (5%) left graduate school without any degrees.

Of the 21 "Rejects":
- 13 (62%) achieved a PhD degree in psychology.
- 2 (10%) achieved an MA degree in psychology.
- 1 (5%) achieved an MD degree.
- 2 (10%) achieved an MA degree in other fields.
- 3 (14%) never entered graduate school.

In summary, of the 245 living assessees:
- 167 (68%) achieved a PhD degree in psychology.
- 34 (14%) achieved an MA degree in psychology.
- 12 (5%) are still graduate students in psychology.
- 6 (2%) achieved an MD degree.
- 8 (3%) achieved PhD or MA degrees in other fields.
- 18 (7%) achieved no known graduate degrees.

---

[a] 1 achieved a PhD in another field.

[b] 1 achieved a PhD in English literature and then returned to psychology.

Using such techniques, questionnaires were received from over 93% of the 245 living *S*s, including 95% of the *S*s who had entered the VA training program in clinical psychology, 96% of the APA members, and 100% of the *S*s who by 1959 had achieved PhD degrees in psychology!

The New Criteria

Academic Attainment

Table 1 summarizes the academic progress of all 245 *S*s by 1957. Approximately 70% of the original *S*s achieved the PhD degree in psychology; interestingly, this figure was not significantly different for a group of 21 *S*s who were not admitted into the VA training program in 1948. It is of some interest to note that 12 *S*s or 5% of the original group still regard themselves as graduate students, even though their degrees are considerably overdue.

Time Required to Achieve the PhD

There was considerable variation in the length of time required to complete work for the PhD degree in psychology, as shown in Table 2. Five years was the general average (mean, median, and mode), but a quarter of the *S*s saw their graduate education prolonged for six years or more.

TABLE 2

Number of Years in Graduate School Prior to the PhD Degree in Psychology

No. Yrs.	No. *S*s	% Sample	Cumulative %
3	7	4%	4%
4	45	25	29
5	61	35	64
6	20	11	75
7	19	10	85
8	6	3	88
9	3	2	90
10	3	2	92
11	2	1	93
12	1	1	94
Still students	12	7	100
Total	179	—	—

Present Work Settings

Table 3 summarizes the major work settings in 1957 of those *S*s who achieved the PhD in psychology. VA and academic settings accounted for over half of this sample, with the remaining *S*s fairly evenly divided among government agencies (non-VA), clinics, and hospitals. Relatively few *S*s worked solely in private practice or in business settings, although almost half of the group reported some private practice or consulting work in addition to their full-time positions.

TABLE 3

1957 Work Settings of the 167 Male Assessees Who Achieved the PhD in Psychology (Major Position Only)

Work setting	No. Persons	% Sample
VA installations	50	30%
Colleges or universities	42	25
U.S. government, state or city agencies or systems	22	13
Clinics, nonhospital	19	12
Hospitals, non-VA	18	11
Private practice	10	6
Business or industrial settings	6	4
Total	167	100%

Satisfaction With Choice of Clinical Psychology

Table 4 indicates the reported degree of satisfaction associated with the choice of a career in clinical psychology for each of several subgroups of *S*s categorized by progress toward the PhD in psychology. On the surface, the finding that only one-half of the total sample would again choose to end up in clinical psychology (if they had their lives to live over again) appears rather discouraging, but any meaningful interpretation of these findings must await a comparison with other professional groups. Even among those receiving the PhD, only three out of five said they would again specialize in clinical psychology.

TABLE 4

SATISFACTION WITH CHOICE OF CLINICAL PSYCHOLOGY BY DEGREE ATTAINED

Academic Progress	If I had my life to live over again (knowing what I now know), I would try to end up in:						
	N	Clinical Psychology	Some Other Field in Psychology	Medicine: Psychiatry	Law or Business	Other	Total
PhDs in psychology	167	60%	12%	16%	2%	11%	100%
MAs in psychology*	26	23	12	19	19	27	100
Students in psychology	12	42		42		17	100
MDs	6			100			
PhDs or MAs in other fields*	7			14	14	71	100
No known degrees*	10	20	10		30	40	100
Total*	228	50%	11%	19%	5%	16%	100%

* Based on incomplete returns.

Relation of Satisfaction to Job Setting

To what extent is satisfaction with one's professional choice determined by situational features in one's present job (as compared to more enduring personality characteristics)? As a first approximation to answering this question, satisfaction ratings were related to present work settings for the 167 PhD's in the total sample. Table 5 presents this analysis. It would appear as if professional satisfaction is only slightly related to work setting.

Relation of Satisfaction to Present Specialization

Perhaps professional satisfaction is related to choice of specialization within the broad field of clinical psychology. Table 6

TABLE 5

SATISFACTION WITH THE FIELD OF CLINICAL PSYCHOLOGY
AS RELATED TO 1957 WORK SETTINGS
(PHDS IN PSYCHOLOGY ONLY)

Work Setting	If I had my life to live over again (knowing what I now know), I would try to end up in:					
	N	Clinical Psychology	Some Other Field in Psychology	Medicine: Psychiatry	All Other Fields	Total
VA installations	50	70%	10%	8%	12%	100%
Colleges or universities	42	64	24	7	5	100
U.S. government, state or city agencies or systems	22	50	9	14	27	100
Clinics, nonhospital	19	63		21	16	100
Hospitals, non-VA	18	44	6	39	11	100
Private practice	10	70	10	20		100
Business or industrial settings	6		16	50	33	100
Total	167	60%	12%	16%	13%	100%

TABLE 6

PROFESSIONAL SPECIALIZATION WITHIN CLINICAL PSYCHOLOGY RELATED TO SATISFACTION WITH THE PROFESSION
(PHDS IN PSYCHOLOGY ONLY)

		If I had my life to live over again (knowing what I now know), I would try to end up in:					
		N	Clinical Psychology	Some Other Field in Psychology	Medicine: Psychiatry	All Other Fields	Totals
At the present time, I consider myself *primarily* a:	Therapist	51	61%	5%	27%	12%	100%
	Teacher	25	72	24	0	4	100
	Researcher	22	45	23	5	27	100
	Diagnostician	22	59	5	31	5	100
	Administrator-Supervisor	46	65	11	7	17	100
If it were not for considerations of training, status, salary, etc. (i.e., in terms of the actual work involved), I would prefer working as a:	Therapist	45	69	0	24	7	100
	Teacher	33	55	15	12	18	100
	Researcher	50	56	22	10	12	100
	Diagnostician	7	86	0	14	0	100
	Administrator-Supervisor	31	58	3	13	26	100
In the years ahead, taking everything into consideration, the odds are that I will devote most of my time to:	Therapist	50	58	0	28	14	100
	Teacher	20	70	20	0	10	100
	Researcher	29	48	24	14	14	100
	Diagnostician	9	56	0	33	11	100
	Administrator-Supervisor	59	68	10	7	15	100

summarizes the findings from three of the questions relating to professional specialization. In general, there appear to be three rather clear-cut areas of specialization for our sample: therapists, academicians (combining teaching and research), and administrators; only a very small percentage of the *S*s claimed to prefer the work of diagnosis or felt they would devote most of their time in the years ahead to diagnostic work.

It would appear as if specialization is only slightly related to satisfaction with clinical psychology; however, there were large and consistent differences between the specialties with respect to which alternative careers seemed most appealing. That is, therapists were more likely than the other groups to state they would enter medicine (psychiatry) if they could begin their lives anew and researchers were more likely to be attracted by other fields in psychology (general experimental, for instance). This would seem to be related to the opportunities to carry out their preferred activities (therapy, research, etc.) in the various professions; the few *S*s who preferred the work of diagnosis over that of other specialties were the most likely to express satisfaction with their professional choice as clinical psychologists.

Specialization as Related to Level of Academic Attainment

Table 7 summarizes the replies of all *S*s with respect to their present, most preferred, and estimated future areas of specialization.

Looking first at column one, it is obvious that in spite of the continuing shortage of psychologists, not all PhDs have found it possible to find a position permitting them to

TABLE 7

SPECIALIZATION IN CLINICAL PSYCHOLOGY AS RELATED TO LEVEL OF ACADEMIC ATTAINMENT

		Level of Academic Attainment				
		PhD in Psychology ($N=167$)	MA in Psychology ($N=19$)	Still Graduate Students ($N=10$)	Degrees in Other Fields ($N=10$)	No Degrees ($N=11$)
At the present time, I consider myself primarily a:	Therapist	31%	6%	20%	38%	36%
	Teacher	15	18	0	25	0
	Researcher	13	18	10	0	0
	Diagnostician	13	35	60	12	0
	Administrator-Supervisor	27	24	10	25	64
	Totals	100%	100%	100%	100%	100%
If it were not for considerations of training, status, salary, etc. (i.e., in terms of the actual work involved), I would prefer working as a:	Therapist	27%	16%	30%	33%	44%
	Teacher	20	16	20	56	0
	Researcher	30	16	20	0	0
	Diagnostician	4	21	30	0	0
	Administrator-Supervisor	19	32	0	11	56
	Totals	100%	100%	100%	100%	100%
Taking everything into consideration, I will probably devote the larger part of my time in the years ahead to:	Therapist	30%	5%	30%	50%	44%
	Teacher	12	21	0	30	0
	Researcher	17	16	10	0	0
	Diagnostician	5	21	50	10	0
	Administrator-Supervisor	35	37	10	10	56
	Totals	100%	100%	100%	100%	100%

perform the work they would most prefer. Thus 13% reported that they were primarily "Diagnosticians" but only 4% preferred that type of work. Only 13% considered themselves primarily "Researchers" but 30% would prefer it. More of the PhDs were doing administration than would prefer it and an even larger proportion anticipated it as their most probable future work. Interestingly enough, the proportions of those doing, preferring, and expecting to continue doing therapy were very close to each other.

While the *N*s for the non-PhD groups are too small to yield stable percentage values, certain of the findings are of interest; e.g., smaller proportions doing and preferring therapy and larger proportions doing and preferring diagnosis.

Scholarly Productivity

As a rough measure of the level of scholarly activity of our *S*s, we searched the indices of *Psychological Abstracts* for the years 1950–1957 inclusive. Omitting unpublished theses but including all other citations, the scholarly output of the total original sample and for the subsample receiving the PhD in psychology is shown in Table 8.

Obviously, this group of psychologists has not contributed much volume—at least as yet—to the current literature of psychology. Approximately half of them had no publications at all during this period, i.e., they did not even publish an article based on their theses. A quarter more have but one publication during the eight-year pe-

TABLE 8

Scholarly Productivity of PhDs and of Total Sample over Eight-Year Period

No. Publications in Psychology	% PhDs (N = 167)	Cumulative %	% Total Sample (N = 245)	Cumulative %
0	47%	47%	58%	58%
1	23	70	20	78
2	11	81	8	86
3 or 4	10	91	7	93
5 or 6	5	96	4	97
7 or more	4	100	3	100

riod, in most cases based on the PhD dissertation. While the scholarly production of other groups of psychologists seems to be roughly comparable to our sample (Dennis, 1954), the facts reported in Table 8 raise serious doubts regarding the effectiveness of the procedures by which this group was presumably selected for research motivation and trained in research methodology.

Intercorrelations of the New Criteria

Although these new criteria are logically distinguishable, it is not likely that they are completely independent. The actual intercorrelations among five of them (for the 1948 sample of 117 assessees) are shown below:

	A	B	C	D	E
A. Academic attainment	—	38	77	79	51
B. Scholarly productivity		—	16	28	05
C. Present identification with clinical psychology			—	70	51
D. APA membership				—	18
E. Combined criterion: Satisfied PhDs vs. all others					—

While all of the criterion intercorrelations are positive, none is so high as to suggest identity and several of them are surprisingly low. From inspection of the matrix, it is obvious that scholarly productivity is relatively independent of the remaining cluster of four correlated criteria.

Correlates of the New Criteria

Eventual Academic Attainment

As earlier criteria of academic performance, Kelly and Fiske (1951) were forced to rely chiefly on ratings by university staff members and scores on an objective content examination in psychology. In spite of the restricted range of talent in this group of PhD candidates, both of these criteria of "caliber" were predicted fairly well by Miller Analogies scores and by the Strong Psychologist Key (correlations in the .40s and .50s). Since over half of the total sample were still working toward their PhDs at the time the earlier report was prepared, it was not possible to provide a definitive analysis of the correlates of eventual academic attainment. Preliminary analyses reported by Kelly and Fiske (1951, pp. 190–192) indicated that this type of "over-all pass-fail" criterion was not likely to be as predictable as criteria reflecting *quality* of graduate academic performance.

Now that most of the original *S*s have either obtained the PhD or have ceased working toward it, we are in a position to examine the correlates of progress toward the PhD. Rather than a simple two-point scale of "PhD vs. no PhD," the 1957 criterion of academic progress consists of a five-point scale, essentially bimodal in distribution. On one end of the scale are those *S*s who had completed their PhD in psychology by 1957; at the other end are those *S*s who had left graduate school without having received such a degree; a middle group (N = 12) is composed of *S*s who were still registered students in departments of psychology in 1957. Each of the two extreme groups were subdivided: the non-PhDs were divided into groups which left graduate school (typically after a year of schooling or less) without any degree in psychology and another group which achieved MA degrees (or the equivalent) in psychology before withdrawing from graduate school. The PhD group was also subdivided into a "fast" group which achieved their degrees in three to five years after entrance into the VA training program and a "slow" group which took six to ten years to complete requirements for the PhD degree.

The resulting scale of academic attainment can be summarized as follows:

1	2	3	4	5
No degree in psychology	MA degree in psychology	Still graduate students in psychology	6–10 yr. PhDs in psychology	3–5 yr. PhDs in psychology

"Predictor" variables from the 1947 assessment included the Miller Analogies Test (MAT), all six scales of Thurstone's Test of Primary Mental Abilities (PMA), all of the present scales of the Strong Vocational Interest Blank (SVIB)—including the Kreidt Psychologist Keys — the six scales of the Allport-Vernon Scale of Values (A-V), the 13 scales of the Guilford-Martin Battery of Personality Inventories (G-M), the nine standard scales of the Minnesota Multiphasic Personality Inventory (MMPI), 11 of the final pooled ratings from the 1947 assessment (1947), and 86 items from the Biographical Inventory (BI). The Pearson product-moment correlates of academic atttainment from these measures are summarized in Table 9, which lists the variables showing correlations above the .05 level of statistical significance for either the 1947 or 1948 samples together with the corresponding values for the other sample.[2]

TABLE 9

CORRELATES OF ACADEMIC ATTAINMENT

Variable	*r* 1947 Sample (N=127)	*r* 1948 Sample (N=117)
Reasoning ability PMA	.26	—[a]
VA clinical psychologist SVIB	.26	.17
Psychologist—Rev. (SVIB)	.25	.17
MAT	.24	.08
Physician SVIB	.22	.03
Reported frequency of expressing emotions (BI)	.22	.02
Rated social adjustment (1947)	.21	—
Social extraversion (G-M)	.20	−.07
Rated intelligence (1947)	.20	—
Rated over-all suitability for clinical psychology (1947)	.20	—
Chemist (SVIB)	.20	.12
Public administrator (SVIB)	.20	−.09
Age at time of assessment (BI)	−.19	.05
Feel authors have better understanding of human nature than do teachers, ministers, lawyers, advertising men, or salesmen (BI)	.19	.13
Reported membership in well-knit gang during adolescence (BI)	.19	−.07
Rated broad (vs. narrow) interests (1947)	.18	—
Social worker (SVIB)	.18	.11
CPA (SVIB)	.18	−.07
Physicist (SVIB)	.17	.41
Banker (SVIB)	−.17	−.07
Depression (MMPI)	−.17	—
Perceived social adjustment immediately precollege (BI)	.17	−.03
Reports associates perceive *S* to be careless about details (BI)	.17	.00
City school superintendent (SVIB)	−.04	−.24
Reported degree of childhood happiness (BI)	.13	−.24
Number of offices held in high school (BI)	.09	.23
Father's educational level (BI)	−.02	−.23
Mother's educational level (BI)	−.03	−.22
S's marital status—single vs. married (BI)	−.06	.22
Reported severity of inferiority feelings in childhood (BI)	.07	.19
Urban vs. rural childhood residence (BI)	−.04	.19
Reported frequency of debilitating periods of discouragement (BI)	−.13	.19
Father's occupational status (BI)	.07	−.19
Cultural and intellectual level of childhood home (BI)	.15	−.18

[a] Indicates that predictor variable was not available for the 1948 sample.

---

[2] Correlations of .17 or higher are significant at the .05 level (.23 or higher at the .01 level) for the 1947 sample ($N = 127$); correlations of .18 or higher are significant at the .05 level (.24 or higher at the .01 level) for the 1948 sample ($N = 117$).

As will be noted, only 23 of nearly 200 potential predictor variables for the 1947 sample (and only 11 of the variables for the 1948 sample) showed significant correlations with this new criterion of academic attainment. And, even in these few instances, the resulting correlations were extremely low—in no case accounting for as much as 10% of the criterion variance. More importantly, however, there were *no* predictor variables which showed significant replication in both of the samples!

Scholarly Productivity

As was reported earlier, the publication record of the two samples of former VA trainees does not appear to have been a distinguished one. For this reason it seems all the more important to be able to locate potential predictors of later productivity. Table 10 lists the variables for the 1947 sample which correlate significantly with this skewed (and rather crude) index of scholarly output, the number of citations in *Psychological Abstracts* during the period 1950–1957.

Some 33 variables correlate with productivity for the 1947 sample, but again none of the correlations account for as much as 10% of the criterion variance. And, none of these variables was replicated for the 1948 sample!

TABLE 10

CORRELATES OF SCHOLARLY PRODUCTIVITY[a]

Variable	r 1947 Sample (N=127)	r 1948 Sample[b] (N=117)
Psychologist—Rev. (SVIB)	.33	.17
Banker (SVIB)	−.29	−.05
% time S desires to devote to research (BI)	.28	−.02
Office manager (SVIB)	−.27	.02
Mathematician (SVIB)	.25	.06
Architect (SVIB)	.24	.03
Physicist (SVIB)	.24	.14
Mortician (SVIB)	−.24	−.10
Parents' religion—Protestant or Catholic vs. Jewish, mixed, or none (BI)	.24	.00
Feel authors have better understanding of human nature than do teachers, lawyers, ministers, advertising men, or salesmen (BI)	.24	.05
VA clinical psychologist (SVIB)	.23	.06
Chemist (SVIB)	.22	.09
Theoretical values (A-V)	.21	−.05
Religious values (A-V)	−.21	−.07
Rated broad (vs. narrow) interests (1947)	.21	—
Artist (SVIB)	.21	.02
Policeman (SVIB)	−.21	.01
Reported amount of cheating on exams in college (BI)	.21	.11
Aesthetic values (A-V)	.20	−.03
Social values (A-V)	.20	.15
Reported strictness of religious training (BI)	−.20	−.06
Veterinarian (SVIB)	−.19	−.09
First decided to become a psychologist early vs. late in life (BI)	−.19	−.02
MAT	.18	.15
Physician (SVIB)	.18	.02
Reported amount of guilt reactions to masturbation (BI)	−.18	.09
Reported regularity of church attendance (BI)	−.18	−.17
Political values (A-V)	−.17	−.14
Rated intelligence (1947)	.17	—
Clinical psychologist (SVIB)	.17	−.04
Engineer (SVIB)	.17	.05
Age at time of assessment (BI)	−.17	−.02
Number of offices held in college (BI)	.17	.03

[a] Criterion: Number of citations in *Psychological Abstracts* (1950–1957).

[b] Additional correlates for the 1948 sample which were not replicated for the 1947 sample are not listed in this table.

APA Membership

Approximately 80% of the original *S*s considered their affiliation with the profession of psychology strong enough to have become members of the American Psychological Association by 1957. In an effort to better understand the characteristics of APA members (as opposed to former graduate students who drifted out of the profession of psychology), all of the variables were related to 1957 APA membership. The criterion utilized for this purpose was simply a two-point scale: *S*s listed in the 1957 edition of the APA *Directory* vs. *S*s not so listed. The correlates of APA affiliation are shown in Table 11.

Only 13 of the variables for the 1947 sample and only 6 of the variables for the

TABLE 11

CORRELATES OF APA AFFILIATION

Variable	r 1947 Sample (N=127)	r 1948 Sample (N=116)
Reported membership in well-knit gang during adolescence (BI)	.27	.02
Age at time of assessment (BI)	−.25	.06
MAT	.25	.13
Reasoning ability (PMA)	.21	—
Pharmacist (SVIB)	.21	−.05
Stated degree of preference for social activities—as opposed to solitude (BI)	.19	.04
Sales manager (SVIB)	.19	−.03
Urban vs. rural childhood residence (BI)	−.18	.12
Reported frequency of expressing emotions (BI)	.18	−.17
Degree of participation in athletics in college (BI)	−.17	−.14
Blames father (rather than mother) for disagreements between parents (BI)	.17	−.16
Numerical ability (PMA)	.17	
VA clinical psychologist (SVIB)	.17	.13
Reported severity of inferiority feelings in childhood (BI)	−.05	.24
Feel authors have better understanding of human nature than do teachers, lawyers, ministers, advertising men, or salesmen (BI)	.12	.24
Number of offices held in high school (BI)	.04	.21
Reported frequency of debilitating periods of discouragement (BI)	−.09	.20
Economic status of childhood home (BI)	−.08	−.20
Number of times family moved during childhood (BI)	.03	.19
Age at time of first part-time job (BI)	−.13	.19

1948 sample correlated significantly with APA membership and none of these were replicated between the two samples.

Present Identification With Clinical Psychology

Since about 30% of the psychologists who originally entered clinical psychology as graduate students were no longer working as clinicians in 1957 and no longer considered themselves *clinical* psychologists, an attempt was made to uncover correlates of eventual specialization in clinical psychology. This criterion, like the preceding one, assumes no particular degree of academic attainment; it lumps as "nonclinicians" *S*s who left the field of psychology entirely, along with PhD psychologists who are now identified with other areas within the profession. Conversely, the group of "clinicians" includes *S*s from all levels of academic attainment. The correlates of professional identification as a clinical psychologist are summarized in Table 12.

Again, as with the preceding analyses, predictor-criterion correlations are quite low and generally nonsignificant in the replication sample. On the other hand, replicated correlates appear for the first time, but the correlations are so low (accounting for about 4% of the criterion variance) that their predictive value is negligible. More importantly, however, they have emerged as two out of hundreds of possible replications and therefore could have occurred on the basis of chance alone. The fact that one variable (urban vs. rural childhood residence) yields a significant correlation in opposite directions for the two samples suggests the plausibility of this interpretation.

Satisfaction With Choice of Clinical Psychology

Approximately 50% of the *S*s stated that they would enter clinical psychology again if they had their lives to live over, with the remainder choosing other more preferred occupational choices (see Table 4). In an attempt to identify earlier correlates of later professional satisfaction, all of the variables were correlated with this dichotomous criterion. Table 13 summarizes this analysis.

Again it is obvious that the correlations are generally quite low, and there emerged no replicated correlates of this satisfaction criterion. Consequently, additional analyses of professional satisfaction were attempted.

TABLE 12

Correlates of Present Identification with Clinical Psychology

Variable	*r* 1947 Sample ($N=127$)	*r* 1948 Sample ($N=117$)
Social extraversion (G-M)	.31	−.06
Urban vs. rural childhood residence (BI)[a]	−.26	.19
Number of "psychosomatic" symptoms reported (BI)	−.25	.10
Preference for diagnostic interviewing—as opposed to therapy or research (BI)	.24	.00
Age at time of first date (BI)	.23	−.09
Reported frequency of expressing emotions (BI)	.23	−.04
Parents' religion—Protestant or Catholic vs. Jewish, mixed, or none (BI)	.22	.08
Reasoning ability (PMA)	.22	—
VA clinical psychologist (SVIB)	.22	.07
Personnel director (SVIB)	.22	.16
Social worker (SVIB)	.22	.14
Special promotion(s) in school (BI)[b]	.21	.20
Feel authors have better understanding of human nature than do teachers, lawyers, ministers, advertising men, or salesmen (BI)[b]	.20	.20
Hypomania (MMPI)	.20	—
Rated broad (vs. narrow) interests (1947)	.20	—
Rated degree of talkativeness (1947)	.20	—
Rated degree of emotionality (1947)	.19	—
MAT	.19	.13
Psychologist—Rev. (SVIB)	.18	.06
Public administrator (SVIB)	.18	−.06
Cultural and intellectual level of childhood home (BI)	.17	−.03
Economic values (A-V)	−.17	.07
Spatial ability (PMA)	.17	—
Depression (MMPI)	−.17	—
Number of offices held in high school (BI)	.00	.28
Social science high school teacher (SVIB)	.08	.21
Occupational level (SVIB)	−.15	−.19
Minister (SVIB)	.09	.18

[a] Statistically significant correlate in opposite directions for the two samples.

[b] Replicated correlate for 1947 and 1948 samples.

TABLE 13

Correlates of Satisfaction with Previous Choice of Clinical Psychology[a]

Variable	*r* 1947 Sample ($N=127$)	*r* 1948 Sample ($N=117$)
MAT	.26	−.02
Rated intelligence (1947)	.26	—
Rated broad (vs. narrow) interests (1947)	.24	—
Reasoning ability (PMA)	.23	—
Rated over-all suitability for clinical psychology (1947)	.21	—
VA clinical psychologist (SVIB)	.21	.00
Reported frequency of debilitating periods of discouragement (BI)	.21	−.02
Reported feelings regarding number of siblings—too few to too many (BI)	.20	.08
Political values (A-V)	−.20	−.09
Depression (G-M)	.18	.11
Cycloid disposition (G-M)	.18	.04
Inferiority feelings (G-M)	.18	.16
Nervousness (G-M)	.18	−.01
Cooperativeness (G-M)	.18	.10
Rated social adjustment (1947)	.18	—
Social worker (SVIB)	.18	.17
Theoretical values (A-V)	.17	−.04
Psychologist—Rev. (SVIB)	.17	−.02
Osteopath (SVIB)	−.17	.00
Marital status—single vs. married (BI)	.03	.25
Age at time of first date (BI)	.15	−.23
Experimental psychologist (SVIB)	.01	−.22
Number of offices held in high school (BI)	−.12	.21
Aesthetic values (A-V)	.06	−.21
Reported difficulty in leaving parents during adolescence (BI)	.05	−.21
Guidance psychologist (SVIB)	.01	.20
Degree of participation in athletics in high school (BI)	.01	.20
Reported amount of guilt feelings in childhood (BI)	.01	.19

[a] Criterion: If I had my life to live over again (knowing what I now know) I would try to end up in clinical psychology vs. all others (Question 9 of 1957 Follow-up Questionnaire. See Appendix B).

Satisfaction With Area of Specialization Within Psychology

Perhaps satisfaction with one's area of specialization within the profession is a less elusive criterion than general occupational satisfaction. To check on this, an index of satisfaction with one's area of specialization was constructed by examining the questions of the follow-up questionnaire relating to present, most preferred, and estimated-future work. Roughly 50% of the *S*s who answered these questions were carrying out activities which they professed to prefer most and simultaneously estimated they would continue with in the years ahead. The other half of the sample either preferred other activities than those they were carrying out or estimated that in the future they would be unable to continue with their most-preferred activity, or both. As in the preceding analyses all variables were correlated with this dichotomous criterion of satisfaction; the findings are summarized in Table 14.

Again, correlations were low and no replications occurred. For this reason, still other analyses were attempted in an effort to find consistent correlates of various aspects of professional satisfaction.[3] In each instance a promising pattern of predictors for the 1947 sample did not replicate for the 1948 group!

TABLE 14

CORRELATES OF SATISFACTION WITH PRESENT AREA OF SPECIALIZATION WITHIN PSYCHOLOGY[a]

Variable	*r* 1947 Sample (*N*=127)	*r* 1948 Sample (*N*=117)
Public administrator (SVIB)	.26	−.11
Nervousness (G-M)	.25	.06
Masculinity-femininity (G-M)	.23	.04
Inferiority feelings (G-M)	.23	.06
Depression (G-M)	.21	−.08
Rated intelligence (1947)	.21	—
Rated over-all suitability for clinical psychology (1947)	.21	—
Reported participation in extracurricular activities in high school (BI)	.21	.05
Reported amount of cheating on exams in college (BI)	−.21	.04
Reported amount of sexual experience during adolescence (BI)	−.21	−.03
Cycloid disposition (G-M)	.20	−.08
Production manager (SVIB)	.19	.15
Reported satisfaction with high school grades (BI)	.18	.00
Reported frequency of debilitating periods of discouragement (BI)	−.18	.03
Masculinity-femininity (SVIB)	.17	−.02
Both parents present in childhood home—vs. one or more missing (BI)	.17	−.05
Father (vs. mother) reported as final authority in childhood discipline (BI)	.17	−.07
Special promotion(s) in school (BI)	.17	.00
Urban vs. rural childhood residence (BI)	.00	.27
City school superintendent (SVIB)	−.08	−.25
Reported feeling regarding number of siblings—too few to too many (BI)	.09	.23
Religious values (A-V)	−.04	.20
Number of times family moved during childhood (BI)	.12	.20
Pharmacist (SVIB)	−.04	−.20
Reported degree of "bookishness" of *S* during adolescence (BI)	.06	.19
Reported regularity of parents' church attendance (BI)	.07	.18

[a] Criterion: Preferred and estimated-future area of specialization both identical with present area of specialization (Questions 6, 7, and 8 of the 1957 Follow-up Questionnaire. See Appendix B).

Academic Success Plus Satisfaction With Professional Choice

Approximately 40% of the total sample had achieved a PhD degree in psychology by 1957 and simultaneously expressed satisfaction with their professional choice (by stating that they would have entered the field of clinical psychology if they could live their lives over again). The other 60% either had not achieved the PhD degree in psychology or felt that they would enter some other field if they could live their lives anew. Since it was felt that such a combined index might be a conceivable criterion upon which to judge initial selection procedures (since ideally a profession might well strive for both "successful" and "satisfied" practitioners), all variables were correlated with this dichotomous multiple criterion. Table 15 summarizes these findings.

---

[3] The analyses of the various indices of satisfaction were all carried out by Antonio Monteiro.

TABLE 15

CORRELATES OF THE COMBINED CRITERIA: ACADEMIC SUCCESS (PHD) PLUS SATISFACTION WITH THE CHOICE OF CLINICAL PSYCHOLOGY

Variable	r 1947 Sample ($N=127$)	r 1948 Sample ($N=117$)
Rated intelligence (1947)	.26	—
VA clinical psychologist (SVIB)	.25	.06
Objectivity (G-M)	.25	.11
MAT	.24	−.05
CPA (SVIB)	.23	−.09
Reasoning ability (PMA)	.22	—
Cooperativeness (G-M)	.21	.07
Memory (PMA)	.20	—
Rated broad (vs. narrow) interests (1947)	.20	—
Social worker (SVIB)[a]	.20	.21
Perceived social adjustment immediately precollege (BI)	.20	.11
Reported frequency of debilitating periods of discouragement (BI)	−.20	.03
Political values (A-V)	−.19	.02
Rated social adjustment (1947)	.19	—
Personnel director (SVIB)[a]	.19	.22
Theoretical values (A-V)	.18	.07
Social extraversion (G-M)	.18	.02
Depression (G-M)	.18	.11
Inferiority feelings (G-M)	.18	.12
Rated over-all suitability for clinical psychology (1947)	.18	—
Veterinarian (SVIB)	−.18	−.01
Reported ability to control emotions (BI)	.18	.04
Spatial ability (PMA)	.17	—
Cycloid disposition (G-M)	.17	.09
Ascendance—Submission (G-M)	.17	.08
Psychologist—Rev. (SVIB)	.17	.04
Number of offices held in college (BI)	.17	.12
Marital status—single vs. married (BI)	.07	.29
Reported amount of difficulty in leaving parents during adolescence (BI)	.00	−.23
Guidance psychologist (SVIB)	.04	.20
Reported amount of guilt feelings in childhood (BI)	−.01	.20
Father's occupational status (BI)	−.02	−.20
Age at time of first date (BI)	.03	−.20
Social science high school teacher (SVIB)	.04	.19
Experimental psychologist (SVIB)	−.02	−.19
Perceived satisfaction of parents regarding *S*'s high school grades (BI)	.09	−.19

[a] Replicated correlate for 1947 and 1948 samples.

Again the correlations are low and the only replicated variables are two highly correlated Strong scales (Social Worker and Personnel Director). As was pointed out earlier, the possibility of this replication emerging on the basis of chance factors alone must be considered.

The Absence of Significant and Consistent Correlates of the New Criteria

In spite of the socially relevant (and reasonably objective) nature of the new criteria, it is clear that only a small proportion of their variance is predicted by the assessment data obtained prior to the *S*s entry to graduate schools. While the findings reported deal primarily with the more objective assessment variables, additional analyses suggest that the pooled ratings of the assessment staff (based on projective and interview data) fare no better as predictors of these new criteria. At first glance, it seems unusually puzzling that these seven real-life criteria appear to be completely uncorrelated with nearly 200 predictor measures covering a wide range of psychological domains. How could this be? Acknowledging the risks of *post hoc* hypothesizing, a few possible explanations merit discussion.

Although many of the new criteria appear to be objective indices of performance, nonetheless they may mask sizable differences in the actual functioning of the *S*s. For example, the standards of performance demanded for the PhD are probably quite varied across universities. Standards of accomplishment in academic and practicum work may vary enormously in schools and agencies with differing goals and methods of training. Thus, a criterion such as attainment of a PhD—while desirably reflecting some uniformity in training and assessed competence—may in fact be remarkably situationally specific. Table 16 presents some evidence for this point of view. Although the small size of the samples for each university make statistical analysis unfeas-

TABLE 16

INTERUNIVERSITY DIFFERENCES IN THE ACADEMIC ATTAINMENT, APA AFFILIATION, AND PROFESSIONAL SATISFACTION OF THEIR GRADUATE STUDENTS

University Code	No. *S*s Assessed	% APA Members in 1957	% Completing PhD in Psychology	% Satisfied with Choice of Clinical Psychology
A	4	75%	50%	25%
B	8	100	75	38
C	5	40	40	0
D	22	96	86	59
E	5	100	80	40
F	12	100	67	58
G	2	100	50	50
H	1	100	100	0
I	1	100	0	100
J	7	57	43	14
K	6	83	83	67
L	9	78	56	44
M	10	80	60[a]	20
N	4	100	100	100
O	7	86	86	71
P	3	67	67	67
Q	4	50	25	0
R	3	100	100	67
S	6	50	67[a]	33
T	9	78	78	89
U	3	100	100	33
V	8	88	75	75
W	6	67	50[a]	50
X	3	100	100	100
Y	4	75	75	50
Z	5	80	60	20
AA	6	100	67[a]	67
BB	10	60	40	0
CC	6	100	83	50
DD	13	92	69	62
EE	5	60	60	40
FF	8	50	50[a]	62
GG	4	100	100[b]	25
HH	8	88	88[a]	38
II	4	100	75[a]	75
MM	1	100	0	0
NN	1	100	100	100
"Rejects"	21	62	62	24
Totals 37 Av. %	245	197 80%	167 68%	113 46%

[a] Includes one *S* who received his PhD at another university.

[b] Includes two *S*s who received their PhDs at other universities.

ible, it is apparent that there are considerable interuniversity differences (for example, University BB vs. University DD), in the eventual outcome of their graduate students on these three criteria. To the extent that these differences exist, no measures of the *S*s alone could ever predict such a criteria effectively! To the extent that our criteria reflect the judgments of other persons (i.e., teachers and supervisors in the case of academic attainment, editors in the case of scholarly productivity), maximal prediction demands the assessment of both those doing the judging as well as those being judged. For a more detailed elaboration of this position, see Stern, Stein, and Bloom (1956).

Even criteria such as perceived satisfaction with psychology as a profession and satisfaction with an area of specialization probably depend in large part upon the kinds of settings in which the *S*s have worked and upon the kinds of experiences which they have encountered in these settings. And again, to the extent that the variance in satisfaction indices are determined by situationally specific variables, to this extent are they unpredictable from measures (no matter how inclusive and accurate) of the *S*s alone.

The possible situationally specific nature of many of the criterion measures appears to be the most reasonable explanation for the general low correlations found in this follow-up study. Nevertheless, there are other possible explanations. It is possible that the range of predictor variables was just not wide enough to include those variables most relevant for the prediction of the eventual criteria; similarly, it is conceivable that the relevant abilities, interests, and attitudes most highly correlated with the new criteria are "late-flowering" and only become measurable *after* the exposure to certain experiences during training (or after a certain number of years of chronological-psychological maturation).

In addition, possible explanations for the lack of significant correlates may lie in the methodology of the study. Because of the essentially dichotomous nature of most of the new criteria and because of the labor-saving advantages of machine analysis of the hundreds of variables under study, no attempt was made to look for nonlinear relationships which may exist between predictor and criterion variables. It is possible that the actual relationships between the two sets of variables may not be as simple as a straight line (i.e., perhaps *S*s in the middle range of some variable are most apt to be highly productive or to attain the PhD while those on the extremes of the variable tend to leave the profession). And, the relationship between predictor and criterion variables might well be even more complicated, involving patterns of two or more variables, demonstrable only by configurational analyses. In many ways it seems more realistic to assume that criteria of the kind with which we were working should bear a more complex relationship with predictor variables than the simple straight line.

Another methodological problem of this study involves the reliability of the criterion indices. In the case of academic attainment, it is fair to assume that there will be no marked changes in this criterion during the ensuing years. In the case of scholarly productivity, however, this is not so obvious. The findings of Girden and Dennis (1954, p. 633) that psychologists reach their maximum productivity rates in their early thirties and that the rate of scholarly output appears to be quite stable over their professional careers, suggest that this criterion might also be a reasonably reliable one (the median age for our *S*s was approximately 35). On the other hand, the measures of satisfaction were obtained on the basis of answers to a questionnaire and it is possible that such answers might fluctuate as a function of present mood.

The above discussion has been concerned solely with the lack of predictively useful correlations between the two sets of variables, the most disconcerting finding of this study. Almost equally disconcerting, however, is the general finding that what appears to be a meaningful pattern of significant (even though weak) predictor variables of each criterion for the 1947 group does not reappear when we examine the 1948 sample.

Why this failure to replicate? There appears to be at least three possible explanations: chance, conditions of assessment, and actual differences in the two samples.

Chance. The tables in this section were constructed by selecting from nearly 200 those variables which correlated significantly at the .05 level or better with the criterion. In most cases, the number of significant correlations was greater than 5% of the total. However, the predictor variables are far from independent, and hence, the number of "significant" correlations to be expected by chance alone is an undetermined value greater than 5%.

Conditions of assessment. The 1947 group came to Ann Arbor for a week-long volunteer assessment program in the summer of 1947 *after* they had been accepted by a graduate school and the VA training program. By contrast the 1948 sample were required to complete the assessment battery at designated centers in the spring of 1948, *before* they had been admitted to graduate school, under the impression that their test scores would (or at least might) be used as criteria of selection by the graduate school of their choice. Actually, this was not the case, since the scores were never reported back to the universities. Nevertheless, it is not unlikely that taking the tests under an apparent selection situation may have resulted in sufficient "faking good" of responses to have resulted in the differences in correlations for the two groups. While this tendency—especially on the part of psychology majors—may have influenced the scores of certain of the inventories, it is not likely that it affected performance on the MAT or responses to the more factual questions of the BI.

Real differences in the two samples. Even though either of the above explanations is sufficient to explain the differences in results for the two samples, we should not overlook the possibility that the groups differed in one or more critical ways which would lead to different real patterns of relationships between the predictor and criterion variables. Thus far we have not identified any such suppressor or modulating variables, but this seems to be a lead worth further investigation. It should be remembered that most of the 1947 group had completed college before entering the military service and many of them had service duties closely related to, if not actually in, clinical psychology. By contrast, the 1948 *S*s, while also mostly veterans, returned from the service to complete their AB or BS degrees. A large portion of them were seniors during the 1947-1948 school year, at which time the VA program represented a very attractive opportunity for graduate and professional training—especially when compared to alternatives then available.

Whatever the correct explanation of the failure to replicate, it is clear that, at the present time, we simply do not have adequate bases for predicting these real-life criteria of professional status. If it is assumed that the lack of consistent correlations is the result of such good selection practice that all accepted applicants are equally likely to succeed on any of several criteria, this would be regarded as a healthy state of affairs. To the authors, however, it seems more likely that our findings reflect the interaction of several factors about which we should not be too complacent:

1. Marked differences among universities with respect to: (*a*) caliber of students accepted for the PhD; (*b*) aptitudes sought in graduate students, e.g., mathematical vs. verbal; (*c*) types of courses offered and required for the PhD; (*d*) relative emphasis in teaching and grading on memory, originality, creativity, rigor, etc.; (*e*) relative emphasis on performance in courses, field work, and thesis; (*f*) nature of and standards for the PhD thesis.

2. The absence of clear-cut and agreed-upon performance criteria for professional employment and promotion. In view of the differences enumerated under No. 1 above, it is probable that "PhD in Psychology," while an objective criterion, has surprisingly little uniformity of meaning from university to university and from specialty to specialty. Add to this the fact that we have no agreed-upon criteria of competence in teaching, diagnosis, therapy, or administration, nor for the quality of publications, and it is hardly surprising that the variance in our new criteria seems to be determined largely by situational rather than intra-individual variables.

PERSONALITY CORRELATES OF SPECIALIZATION IN PSYCHOLOGY[4]

Although our efforts to find consistent correlates of performance and satisfaction *within* the field of psychology were not successful, the early work of Strong (1943) and later of Nachmann (1957) had already demonstrated marked differences *between* various vocational groups. In the past few years a number of studies have focused on the personality correlates associated with different professions, comparing groups whose work activities appeared quite dissimilar (i.e., accountants and creative writers [Segal, 1953]). Since many professions, including psychology, are extremely broad in function and their members engage in quite different kinds of work activities, it appears reasonable to assume that persons who choose (or are chosen) to perform similar types of work—regardless of the profession with which they are identified—will have certain important personality characteristics in common, and thus may well have shared some critical developmental experiences. Conversely, within any broad occupational group, members who perform different types of work activities might be expected to differ from each other in a manner congruent with their differing intravocational preferences. To test this assumption within the framework of psychology, samples of *S*s representing major fields of specialization were compared on variables from the objective measures common to the 1947 and 1948 assessments.

Originally five areas of specialization were considered: therapy, diagnosis, research, teaching, and administration. However, responses to the 1957 follow-up questionnaire revealed few *S*s who intended to devote their professional lives to the practice of diagnosis and few *S*s who were either "pure" teachers or "pure" researchers (rather than some combination of the two activities). For this reason, it was judged most practical to consider three major groups of specialists: therapists, academicians (representing the combined activities of teaching and research), and administrators. In a sense, these three subfields can be seen as reflecting a logical division among many professions—the practitioner, the academician, and the administrator being in no way unique to the field of psychology. For this reason, it is hoped that some of the correlates of specialization among psychologists might be relevant to choice of specialization among other professional groups.

---

[4] This section is based on the junior author's doctoral dissertation (Goldberg, 1958).

In order to increase the probable homogeneity of the specialist criterion groups, the samples of each were selected from among the total 1947 and 1948 samples to include only persons who (*a*) reported they preferred the work of their specialty to that of any other and (*b*) who simultaneously stated that they believed they would devote the largest proportion of their future professional careers to the same work. Using these criteria, 20 relatively "pure" representatives of each of the three subfields were selected. Table 17 shows a comparison of the three samples on selected variables which described their status in 1957.

TABLE 17

A COMPARISON OF THE THREE SPECIALIZATION SAMPLES ON SOME VARIABLES INDICATING THEIR STATUS IN 1957

	Therapists ($N=20$)	Academicians ($N=20$)	Administrators ($N=20$)
Assessment year:			
1947	10	8	15
1948	10	12	5
Present age:			
Mean age	36.5	34.7	35.0
Range	32–44	31–44	31–41
Geographical distribution:			
New York State	4	5	0
New England and East Coast	2	8	8
Midwest	9	4	7
South	1	1	4
Southwest and Rocky Mountain States	2	0	1
West Coast	2	2	0
Present field in psychology:			
Clinical psychology	20	14	20
Experimental or social psychology	0	6	0
Work setting:			
VA installations	4	4	5
Colleges or universities	0	14	2
Government agencies (non-VA)	0	2	10
Clinics (non-hospital)	8	0	0
Private practice	8	0	0
Hospitals (non-VA)	0	0	3
Satisfaction with professional choice ("If I had my life to live over again, I would choose"):			
Clinical psychology again	16	11	13
Some other field in psychology	0	7	2
Medicine (psychiatry)	4	0	1
Law or business	0	0	2
Other	0	2	2
Number of years past PhD (to 1957):			
Mean number of years	4.5	5.1	4.7
Range	2–7	3–7	2–6

In an effort to isolate variables which would differentiate therapists, academicians, and administrators (with the hope of thereby being better able to understand the basis for their intravocational choices), *S*s from the three subfields were compared on all of the variables from the measures selected for this study. Variables which differentiated between any two of the three groups at the .05 level of statistical significance (chi square, corrected for continuity) were retained for further analyses.

All of the variables which differentiated any pair of the subfields were intercorrelated. Since many of the variables were highly correlated, a series of factor analyses was carried out to identify the underlying factors which could account for these relationships. The centroid solutions were rotated by the Varimax method (Kaiser, 1958) and factor estimate scores were computed in order to permit comparison of the three groups on each factor.

The Factors Underlying Professional Specialization

Forty-eight variables were found to differentiate significantly among therapists, academicians, and administrators, and nine factors appeared to underlie these variables. The nine rotated factors were named and interpreted as follows:

Factor I: Childhood rebelliousness.[5] This factor appeared to tap the amount of conflict and disturbance within the family group. On one pole of this factor were loaded variables such as reported friction with the father, discipline authority always one parent (as compared to shared responsibility for discipline), and childhood feelings of inferiority; while on the other pole were such variables as reported understanding and permissiveness of the parents during the *S*'s adolescence, and over-all childhood happiness. This factor markedly differentiated academicians (rebellious) from administrators (nonrebellious), and this difference can be interpreted in terms of the differing amounts of freedom from authority intervention which characterizes the work of the academician and the administrator.

Factor II: Intrapsychic conflict. This factor was our nearest approximation to the "neuroticism" factor often uncovered in factor analytic studies. Highly loaded on one pole of this factor were such variables as reported childhood guilt feelings, vivid and frequent childhood fantasy, cutting classes frequently in college, eldest child in birth order, and a number of vocational preferences interpreted in terms of an underlying dimension of creative, yet neurotic, fantasy life. This factor differentiated academicians (high conflict) from administrators (low conflict), and this can be interpreted in terms of the differences in the use of fantasy in academic work as opposed to the reality orientation of administrative work.

Factor III: Social conventionality. This factor appeared to tap a constellation of variables which may be thought of as characterizing "the all-American boy." Highly loaded on one pole of this factor were such variables as Protestant or Catholic religious backgrounds (as opposed to Jewish families, mixed marriages, or agnostic parents), regular church attendance, participation in college athletics, and political values (A-V). On the other pole were such variables as theoretical values (A-V) and high MAT scores. This factor—which can be interpreted in terms of the conventionality of the basic values being tapped—differentiated administrators (highly conventional) from both other groups. Such a finding seemed congruent with a view of administrative activities as most approximating the work of business personnel—who, in turn, might be expected to reflect the most conventional values of our society.

Factor IV: Curiosity. This factor appeared to tap either some aspect of curiosity about the world or, perhaps, a particular kind of intellectual functioning. Loaded on one pole were variables such as difficulty with mathematics and science in high school and sex information received late in life; MAT score loaded on the other pole. The factor differentiated academicians (most curious) from administrators (least curious) and seemed congruent with the differing opportunities for satisfying curiosity in the two kinds of work. On the other hand, the reason why therapists failed to present a more homogeneous appearance on this factor was not understood.

Factor V: Exploitativeness. This factor, similar in many ways to factors found in other studies, was characterized by preferences for leadership and administrative positions and by high political and economic values (A-V). The term "go-getter" seems descriptive of individuals at one pole of this factor. Administrators were markedly more exploitative than either of the other two groups; and this finding can be interpreted in terms of the greater opportunities for leadership inherent in administrative work.

---

[5] The first six factors include variables from the BI. It is important to realize that these variables, based on the *S*'s evaluation of his childhood state while looking back at it from the vantage point of young adulthood, make any exact interpretation of their meaning hazardous because of their extreme subjectivity. For these variables the question can be raised as to whether differences in responses reflect genuine differences perceived in childhood, differences perceived only when looking back upon childhood, or merely differences in the "honesty" of the respondent to admit socially unattractive personality attributes (response set). Consequently, although the factors are interpreted here in terms of actual developmental differences, the reader is cautioned that other interpretations are also plausible.

Factor VI: Vocational indecisiveness. This factor was difficult to interpret since few variables had high loadings on it. On one pole were such variables as deciding to become a psychologist late in college and preferences for leadership positions and for diagnostic interviewing; on the other pole was participation in college athletics. Administrators (least vocationally decisive) were differentiated from both academicians and therapists.

Factor VII: Scientific values and interests. This factor differentiated *S*s whose Strong interest patterns approximated those of scientists in many fields (physics, chemistry, engineering, mathematics, and experimental psychology) from *S*s whose interests more closely resembled sales personnel (sales managers, real estate and life insurance salesmen). Interestingly, reported difficulty with mathematics and science in high school loaded highly on the sales pole of this factor, indicating an interrelation on this factor between abilities and interests. This factor markedly differentiated academicians (highly scientific) from administrators, a finding congruent with their later choice of professional specialization.

Factor VIII: Creativity. This factor differentiated *S*s whose Strong interest patterns resembled persons in artistic fields (artist, architects, author-journalists) from *S*s more interested in business detail. Highly loaded on the artistic pole of this factor were theoretical values (A-V) and MAT score. Academicians appeared markedly more creative in interests than did administrators, and these values appear consonant with the creative character of much of the work in the academic world.

Factor IX: Altruism. The social service scales on the SVIB defined this factor, which was expected to differentiate therapists from both other groups. Although therapists were slightly more similar in interests to persons in other social service fields than were academicians or administrators, the differences were not great. Moreover, therapists did not seem at all homogeneous on this factor.

Table 18 contrasts the three groups on all nine factors. A rotated factor matrix showing the loadings of the significant variables on all factors can be found in Appendix C.

TABLE 18

INTERGROUP MEAN DIFFERENCES BETWEEN THERAPISTS, ACADEMICIANS, AND ADMINISTRATORS ON THE NINE FACTORS[a]

Factors	Specialist Groups		
	Academicians	Therapists	Administrators
Childhood rebelliousness	high[b]	med.[b]	low[b]
Intrapsychic conflict	high	med.	low
Social conventionality	low	low	high
Curiosity	high	med.	low
Exploitativeness	low	low	high
Vocational indecisiveness	low	low	high
Scientific (Strong)	high	low	low
Creativity (Strong)	high	med.	low
Altruism (Strong)	low-med.	high-med.	med.

[a] Group differences are expressed as high, medium, or low on each factor.

[b] All high vs. low differences significant: $P < .01$. All high vs. med. and med. vs. low differences significant: $P < .05$. Low-med. vs. high-med. difference significant: $P < .05$. All other differences nonsignificant (Duncan's Multiple Range Test).

Personality Descriptions of the Modal Therapist, Academician, and Administrator

The following is a summary description of the three subfields in terms of the variables which differentiated them. All statements are to be regarded as relative only to the three groups under comparison—and not necessarily to "people in general."

Academicians. The combined academician sample turned out to present a relatively homogeneous personality picture—in spite of the fact that it contained a number of individuals who were presently not working in university settings. It would thus appear that our conceptualization of an academically oriented group within the larger professional field was a useful formulation.

The early childhood of the academician would appear to be characterized by a considerable amount of intrapsychic disturbance, perhaps brought about in part by a rather dominant mother (with whom the child was enveloped in ambivalent friction) and a more taciturn easy-going father. At any rate, academicians reported having had considerable guilt feelings as children, a good deal of flight into fantasy, and more than average guilt reactions to masturbation. They were likely to experience feelings of inferiority and in general to view themselves as somewhat unhappy during childhood. Moreover, they were not likely to be leaders of their peers and their relationships with other children may well have caused them some anguish. Not being athletically inclined, they apparently did not fit easily into the group activities of childhood.

On the other hand, their relative isolation and solitude as youths—combined with their need to rely on their own fantasy and creative imagination for stimulation—appears to coincide with their later choice of work. The working activities of an academician can be viewed as relatively solitary and creative, in contrast to the interpersonal and organizational activities of many other fields. The introspective orientations of academicians as children seemed to have paved the way for continued activities of this nature.

Moreover, the old observation that less popular children often compensate for their lack of peer attention by turning their energies to scholastic pursuits seems to hold here. Academicians appeared to have been curious as youths and to have done well in school. They were probably more serious students throughout their schooling, and they decided on a specific professional career early in their college lives. By this time they had rejected many of the conventional values of society in favor of ones more suitable to their introspective orientations.

One can see the prospective academician in college as bright, theoretically oriented, widely read, and not socially active. He has rejected religion—at least in its conventional forms—and he is neither exploitative nor pecuniary. He sees the world as a relatively unkind place and is determined to do something about it. More importantly, perhaps, he has found an outlet in books and research for many impulses which he is not adept at expressing interpersonally.

Upon entering graduate school, the prospective academician is still likely to be a serious student and to be more concerned with ideas themselves than with utilizing them to manipulate other persons. His interests approximate more nearly those of scientists—perhaps in all fields—and artistic or creative persons in many areas. His values are diametrically opposed, in many ways, to individuals in the business world. In a word, he is *not* the "organization man."

Administrators. Since administrators and academicians in the field of clinical psychology both chose the same profession in which to spend their lives, they must have shared a certain commonality of early developmental experiences and later interests and values by the time they entered graduate school. Nevertheless, on many dimensions, administrators appeared considerably different. They, too, formed a remarkably homogeneous group, in spite of the fact that they differed with respect to the kinds of settings in which they performed their administrative functions.

If the academician could be characterized as an introvert, the administrator emerged as an extrovert. Often the youngest or only child from a family where the parents shared the same Protestant or Catholic faith, the prospective administrator saw the authority for his discipline shared by both his mother and his father and experienced little friction with either parent. Of all our groups, his was the happiest childhood—or so it would appear from what he reported of it. He was least likely to report recourse to vivid or frequent fantasy; he was least often plagued with feeling of guilt or inferiority; he perceived his parents as most understanding and permissive.

Perhaps even more significant, his relationships with his peers were reported as considerably more pleasant. He was likely to have been a member of a childhood gang, and as often as not was its leader. In high school and college he participated in athletics and was often elected to class and club offices. He was more likely to have some trouble with mathematics and science and was probably not as serious a student, in general, as the prospective academician. Perhaps he was less curious—he learned about sex a little later in life—and was more interested in social activities.

In college his interests and values were more conventional ones, shared—one might guess—by most of his college peers. He was religious, in the sense that he attended church more regularly and considered himself a Protestant or Catholic. He decided on his professional career later in college, perhaps only after some thought of entering the business world. In general, though he was not a scholar, he attended classes faithfully (like church, perhaps), and no doubt achieved fairly high grades. On the other hand, he scored lower on the MAT and was less likely to turn in an outstanding academic performance.

As he entered graduate school, when the administrator thought of future specialization he was more likely to prefer diagnostic interviewing (surely *not* marital counseling). More than academicians or therapists, he wanted to be a *leader* in the field, a departmental *chairman,* or a *director* of some project. The interpersonal orientation so characteristic of him during his developmental years was carried over to his choice of specialty: he would still be a "member of the gang," still be a "leader."

Since every person shares many of the same human attributes, the above descriptions of academicians and administrators can best be described as caricatures of the differences that were found to exist between them. Obviously, most of the individuals in each group would not fit these abstract molds perfectly. However, one can estimate the extent to which these differences, taken in combined form, actually did differentiate the academician from the administrator samples by combining the factor estimate scores over the eight factors which individually differentiated the two groups. When this was carried out, the degree of overlap between academicians and administrators was found to be minimal. Splitting the pooled scale at the median score, only two academicians overlapped on the administrator half of the distribution, while no administrators were misclassified.[6] Therapists, on the other hand, appeared to be normally distributed on this combined dimension.

Therapists. In the original selection of *S*s for this study, the therapist group appeared to be the purest sample of persons representing any of the areas of specialization; consequently, we expected them to present a more homogeneous appearance on the variables considered in the study than either of the other two subfields. In fact, quite the contrary occurred!

This failure to identify more adequately the unique characteristics of therapists presents one of the most enigmatic findings of this study. The working activities of therapists—regardless of their therapeutic orientations—seems so similar that one might expect their occupational choices to be equally similarly motivated and their developmental backgrounds to present a relatively homogeneous appearance. Our findings indicate this is not the case.

Our picture of the therapist, as it emerges from those few variables which discriminated the group from the others, is spotty and uneven—without the consistencies we found in the case of the academicians and the administrators. The therapist was more likely the eldest child in a Jewish family. He reported considerable friction with his father and professed a deep attachment to his mother, suggesting possible failure to resolve the theoretical Oedipus conflict. If we accept this type of formulation of the therapist's dynamics, it might explain his reported greater involvement in vivid and frequent fantasy. As a child, the therapist was likely to experience feelings of inferiority and to share with the academician a disinclination toward athletics.

In high school, although his values are

---

[6] Obviously, a replication is necessary in order to ascertain the loss in discrimination resulting from the removal of chance occurrences.

likely to have been theoretical ones, he is likely to have had some difficulty with mathematics and science. By the time he entered graduate school he had already tentatively decided on doing therapy and was especially intrigued with marital counseling; being the inventor of a personality test or the author of a psychological theory also had some appeal. In interests, he resembled personnel directors, public administrators, and—to a lesser extent—pharmacists; he appeared relatively dissimilar to architects. His interests were rather close to those of guidance and industrial psychologists.

Probably as significant as these variables which discriminated therapists from the other groups was the fact that most of the variables considered in this study failed to differentiate therapists at all. In general, therapists can be found evenly scattered throughout the entire range of most of the variables in our analyses, indicating that this group is an extremely heterogeneous one.

A Small Replication

In view of the failure to find replications of significant relationships between predictor and criterion variables (as reported earlier) it seemed appropriate to question the degree of generality which may be attributed to the findings reported here. Since the specialist groups were selected as the "purest" cases from the combined 1947 and 1948 samples, it was not possible to replicate these analyses. However, as a check on the probable generality of the findings, a test was made of their applicability to 10 new *S*s selected to approximate the criteria of professional specialization used; in comparison with the 60 *S*s on whom this study was based, they were less "pure" representatives of the three areas of specialization within clinical psychology. Three therapists, four academicians, and three administrators composed this new group.

Factor estimate scores were computed for each of these 10 new *S*s on the basis of their test scores and responses to the BI, and then these factor scores were pooled to get estimates of the similarity of these 10 *S*s to the characteristics of the previous therapists, academicians, and administrators.

The resultant findings were very encouraging. All four of the persons in the new sample of academicians achieved factor scores well into the range of the original academicians; moreover, two of the three administrators achieved factor scores in the range of the original administrator sample. If we were to have predicted the specialization of these seven *S*s, we would only have misclassified one (and his combined factor score was quite close to the median score of the original 60 *S*s). The three therapists, as expected, were less homogeneous; however, two of the three achieved scores far into the therapist portion of the combined scale which differentiated the original 20 therapists. These findings cannot be considered as a replication of the original study, but the congruence of the personality characteristics of 10 new *S*s to the original sample certainly suggests the possibility of wider generality of these findings.

SUMMARY

Two hundred and forty-five *S*s, assessed in the VA Selection Research Project in 1947 and 1948 just prior to their entry into graduate school in psychology, were followed up by a mail questionnaire in 1957. Responses to the questionnaire yielded information concerning their occupational and educational histories since graduate school, their present areas of specialization within the profession, and their present satisfaction with their previous vocational choices and their choices regarding specialization. An index of their scholarly productivity was obtained by a search of *Psychological Abstracts*.

Approximately 70% of the original *S*s had achieved a PhD in psychology by 1957 while 5% still considered themselves graduate students; the modal number of years taken to obtain a PhD was five. Roughly 50% of the total sample stated that they would choose some vocation other than clinical psychology, if they could live their lives over again, while approximately 40% of those *S*s who had achieved the PhD in

psychology expressed this form of dissatisfaction with their previous professional choice. Satisfaction appeared to be related—only in small part—to present work setting and present area of specialization within psychology. The scholarly productivity of the total sample seemed unimpressive, with nearly one-half of the *S*s having no publications in psychology and another quarter having only one such publication (typically based on a PhD dissertation) during an eight-year period.

Nearly 200 of the earlier assessment variables, mostly from objective tests and a Biographical Inventory, were correlated with criterion measures of academic attainment, scholarly productivity, American Psychological Association affiliation, present identification with clinical psychology, satisfaction with previous professional choice, satisfaction with choice of specialization, and a combined criterion of academic success and simultaneous satisfaction with professional choice. In general, predictor-criterion intercorrelations were low, in no case accounting for more than 10% of the criterion variance. Moreover, for each criterion a set of significant correlations for the 1947 sample disappeared upon replication with the 1948 sample. Possible explanations for these findings were discussed.

In another part of the study, 20 relatively pure representatives of each of three subfields within psychology were selected on the basis of questionnaire answers concerning their present, most preferred, and estimated-future work activities. These therapists, academicians (combining the activities of teaching and research), and administrators were then compared on the nearly 200 objective variables from the earlier assessments. All of the variables which significantly differentiated any two of the three groups were then factor analyzed, and the groups were then compared on each of the nine factors found. The factors were described, and modal personality characterizations for each of the three groups were presented. A small replication was considered successful. In general, the findings from this part of the study suggest that prospective academicians and administrators—and to a lesser extent therapists—in the field of psychology have differential interests, abilities, and values upon admission to graduate school and that at least some of these personality differences are related to marked differences in their perceptions of their early developmental backgrounds. These findings, if fully replicated, may apply to other professions whose work activities parallel those in psychology.

REFERENCES

Dennis, W. Productivity among American psychologists. *Amer. Psychologist,* 1954, **9**, 191–194.

Girden, E., & Dennis, W. Publication trends in American psychology—a five-year extrapolation. *Amer. Psychologist,* 1954, **9**, 632–635.

Goldberg, L. R. Personality development and vocational choice: A study of therapists, academicians and administrators in clinical psychology. Unpublished PhD dissertation, Univer. of Michigan, 1958.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, **23**, 187–200.

Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. Michigan Press, 1951.

Nachmann, B. Childhood experience and vocational choice: A study of lawyers, dentists, and social workers. Unpublished PhD dissertation, Univer. of Michigan, 1957.

O.S.S. Assessment Staff. *Assessment of men.* New York: Rinehart, 1948.

Segal, S. J. The role of personality factors in vocational choice: A study of accountants and creative writers. Unpublished PhD dissertation, Univer. of Michigan, 1953.

Stern, G. G., Stein, M. I., & Bloom, B. S. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

Strong, E. K., Jr. *Vocational interests of men and women.* Stanford: Stanford Univer. Press, 1943.

(Early publication received July 13, 1959)

APPENDIX A

MAJOR FINDINGS OF THE ORIGINAL KELLY AND FISKE STUDY*

A. Regarding the Subjects and Their Training

1. Graduate students in psychology participating in the VA program are remarkably similar to other graduate students in psychology not majoring in the clinical area. Both are highly selected groups of students, not only in intelligence but also in terms of their measured social and emotional adjustment. Both show a pattern of interests and values characteristic of professional and scientific persons. The VA students differ slightly from the others primarily on the basis of interests more characteristic of the professions which involve contact with people, social welfare, and persuasive activities; the interests of the non-clinical students are somewhat more similar to professions which are primarily concerned with things and ideas.

2. Groups of trainees enrolled at different training institutions show significant differences on measures of ability and achievement.

3. Trainees report wide differences in emphases of training programs at different universities and in their clinical experiences in field training.

4. Ninety-five of the 128 P-1 trainees assessed and entering the four-year program in 1947 will probably complete their training in the VA program. For this group, then, the total attrition is about 20%. However, at least seven more of the original group are known to be completing their PhDs in psychology, some with clinical specialization but not as VA trainees.

5. About one-third of the trainees studied report having received some personal therapy during their training.

6. The Strong Vocational Interest patterns of young clinical psychologists were found to be much more similar to those of young psychiatrists than to those of young physicians. The psychiatrists and clinical psychologists were much more similar in pattern of interests than were psychiatrists and other physicians.

B. Regarding Criteria of Professional Success

1. There is no satisfactory single criterion of success in training for or practicing clinical psychology. University staff members and supervising clinical psychologists show wide individual differences in their conceptions of successful academic performance and of professional skills.

2. There are three general components of success in clinical psychology: (*a*) intellectual accomplishments (e.g., academic achievement and research productivity), (*b*) the clinical skills of diagnosis and therapy, and (*c*) skills in social relations. In general, judges agree much better on the first than on the latter two components.

3. Supervisors' ratings do not adequately differentiate the various job functions of the clinical psychologists.

4. Ratings of clinical competences appear to be as much a function of the role of the rater (e.g., teacher, supervisor, colleague) as of the person being rated.

5. No completely acceptable objective measures of competence in the purported functions of the clinical psychologist are available, but certain promising techniques for evaluating clinical skills were developed within this project. However, further developmental work on these measures is necessary before they can be used in the routine evaluation of clinical competence.

In summary, we may say that there are several varied and alternative criteria of success in clinical psychology. However, the rated criteria showed sufficiently high intercorrelations to identify a common core of what might be called "perceived clinical competence."

C. Regarding the Prediction of Success in Clinical Psychology (for the 1947 subjects)

1. Accuracy of prediction of success in clinical psychology is a function of the job component being predicted.

2. The intellectual aspects of success in graduate training and also the general aspect of over-all clinical competence can be predicted surprisingly well in view of the fact that the group of students entering training is already highly selected. Validities for these criteria range from .35 to .60 even when the range of talent is further restricted to those completing this doctoral program. Furthermore, the validities are based on a national sample; known inter-institution and inter-installation differences in caliber of students and emphases of training presumably lower the obtained relationships.

3. Prediction of relative standing within the group completing professional training is generally more accurate than the prediction of outcome of training (Success/Failure). This is probably due to (*a*) the diversity of causes for leaving the program: e.g., academic failure, health, preference for other fields, etc., and to (*b*) differences in the criteria applied by various universities for continuing in the program.

4. The more clinical aspects of professional competence, being less tangible, and therefore not as readily evaluated, are less predictable than the intellectual component of professional success. Each of several criteria of diagnostic and therapeutic competence is predicted by one or more predictors, but different criteria of these skills tend to be predicted by different predictors.

5. Those aspects of professional success involving social skills are also predictable beyond chance

---

* This section is quoted directly from Kelly and Fiske (1951, pp. 193–197).

but because measures of these skills are in part a function of the persons evaluating them, they tend to be less predictable than intellectual aspects of professional success.

6. For the group of trainees who entered training in 1948, the validities of both objective test scores and assessment ratings based on a modified assessment program were found to be much lower than for the group entering in 1947. These lower validities may be a function of any one or any combination of the following factors:

(a) Test data and other materials were obtained for the 1948 group before action was taken on their applications for admission to graduate school.

(b) The 1948 group is younger, intellectually more able, and more homogeneous with respect to both age and intellectual ability.

(c) The 1948 group had been in training only two and one-half years at the time criterion measures were collected.

D. Regarding the Relative Efficiency of Various Predictors

1. The potential validity of all predictors is apparently limited by inadequacies in available criterion measures. All criterion measures used and considered are more or less fallible because of one or more of the following characteristics. They may (*a*) have low reliability; (*b*) be too general or too specific; (*c*) be lacking in validity, i.e., not relevant to actual competence; (*d*) be too provincial, i.e., reflect only the degree to which professional behavior conforms to that expected in a subculture.

2. Each of the several criterion measures can be predicted about equally well by any one of several techniques or procedures. This is especially so for the criteria of intellectual performance (e.g., rated Academic Performance, scores on Content Examination and rated Research Competence).

3. In general, only a small proportion of the objective test scores were found to correlate significantly with the criterion measures. However, one or two scores from standard psychological tests predict most of the criterion measures about as well as the best of the assessment ratings by professional staff members. Although assessment ratings and objective test scores were found to be about equal in predictive efficiency, objective tests are, of course, much more economical in terms of time and cost. The most generally useful of the objective tests were the Miller Analogies Test and the Strong Vocational Interest Blank. Scores on one or both of these tests predicted most of the several criterion measures with fair efficiency.

4. With respect to the descriptive, evaluative and predictive ratings made by assessment staff members, it was found that: (*a*) predictive ratings show fair to high interjudge reliability when based on the same materials; (*b*) the relationship between the accuracy of assessment prediction and the amount of materials on which the predictions are based is markedly non-linear. In general, assessment predictions based on the credential file plus the objective test profile tend to be almost as accurate as those based on more materials including an autobiography, projective tests, interviews, and situation tests; (*c*) the findings do not reveal any clear-cut superiority in the validity of pooled ratings made in a "staff conference" or of arithmetical ratings over those of the individual staff members; (*d*) self-ratings by the candidates have some predictive value for criteria of intellectual success but show no relationship to other criteria of professional success; (*e*) judgments made by assessment teammates and judgments by staff members on the basis of observing the candidates only in situation tests both have some predictive value; in general they are about as valid as staff ratings based on the credential file alone; (*f*) predictions of success in clinical psychology made on the basis of single projective techniques tend to have very low correlations with any of the criterion measures. The same is true for predictions based on an integrative study of the protocols for the four projective techniques.

5. Predictions of success in the training program made by university staff members on the basis of procedures conventionally used by departments in the selection of students tended to correlate as well with criteria of academic performance and research competence as any assessment predictions. This was true even where the candidate had not been seen by the university staff member at the time of making the predictions. However, the generality of this finding is limited by the fact that university staff predictions were available for only half of the subjects, and criterion measures for this subgroup were more predictable than those for the total group of subjects.

E. Other Findings

1. Qualitative assessment evaluations for 20 trainees were clinically matched with qualitative evaluations of on-the-job functioning made three and one-half years later with an accuracy of 72% where the chance value was 50%. The median phi coefficient for individual cases was .44. The same 20 cases were matched objectively, on the basis of the profiles of personality ratings by teammates at the time of assessment and profiles of ratings made on the same variables by colleagues three and one-half years later, with an accuracy of 74%. The median phi coefficient was .60. Both of these matchings are significantly better than chance.

2. Ratings on all personality variables by staff and teammates at the time of assessment tend to show low correlations with ratings on the same variables made by colleagues three and one-half years later. These correlations are lower than those between the final staff predictive ratings at the time of assessment and criterion measures obtained three and one-half years later. Whether these data indicate a lack of consistency of personality among

the subjects over the period of graduate study reflect different frames of reference for different raters, or merely reflect the relative atypicality of the behavior displayed by candidates at the assessment, is not known.

3. None of the rated criterion measures of clinical competence or scores on the objective Content Examination were found to differentiate the skills of trainees with three and one-half as compared with two and one-half years of training. This finding may be the result of the latter group having been more highly selected or better trained. However, a number of lines of converging evidence lead us to question whether present advanced training procedures, or additional years of experience lead to any currently measurable improvement in clinical skills. This may be a function of the nature of the training and/or experience, or it may merely reflect the inadequacy of available criterion measures of clinical skills.

APPENDIX B

Veterans Administration Selection Research Project, 1027 E. Huron Street,
Ann Arbor, Michigan

1957 Preliminary Follow-up Questionnaire

Code number______

(In order to preserve complete anonymity, please do not sign your name on this form.)

1. Graduate school(s) attended, dates of entry and leaving, and degree(s) received:

School	Dates	Degree(s)

2. Field of graduate specialization: ________________
(if other than clinical psychology)

3. Reason for change of field from clinical psychology (if applicable): ________________

4. Reason for leaving graduate school prior to Ph.D. (if applicable): ________________

5. Please list the positions you have held during the past seven years, starting at the present and working backward in time. List any part-time work performed since leaving graduate school (including private practice, if you had one).

IN THE CASE OF WORK RELATING TO THE FIELD OF PSYCHOLOGY, please try to estimate—as accurately as possible—the percentage of time you spent in each of the following roles: Administration or supervision, diagnosis, research, teaching, and therapy. In cases where your job function changed over the years on one job, please indicate such shifts below.

IN THE CASE OF PERSONS WORKING OUTSIDE THE FIELD OF PSYCHOLOGY, please cross out the listed job functions and substitute ones more appropriate to your work.

				Percentage of time spent in:				
Title, or Name of Position (e.g., Staff Psychologist,	Employer V.A. Hospital,	Location Downey, Ill.	Dates '52-55)	Adminis-tration-Super-vision %	Diagnosis %	Research %	Teaching %	Therapy %

(The answers to questions on this page are requested when they are applicable.)

6. At the present time, I consider myself primarily a:
(Rank order your answers from 1 (MOST) to 5 (LEAST); indicate ties by duplicate ranks.)

______Therapist
______Teacher
______Researcher
______Diagnostician
______Administrator-Supervisor

7. If it were not for considerations of training, status, salary, etc. (i.e., in terms of the actual work involved), I would prefer working as a:
(Rank as above)

______Therapist
______Teacher
______Researcher
______Diagnostician
______Administrator-Supervisor

8. In the years ahead, taking everything into consideration, the odds are that I will devote most of my time to:
(Rank as above)

______Therapy
______Teaching
______Research
______Diagnosis
______Administration-Supervision

9. If I had my life to live over again (knowing what I now know), I would try to end up in: (check one)

______clinical psychology
______some other field in psychology
______law or business
______medicine
______other (please specify): ______________________________

10. In the years ahead, I intend to work in the following kind(s) of setting: (check one or more)

______My home or private office
______A clinic
______An N.P. hospital
______A general medical hospital
______A University, in the Dept. of Psychology
______A University, in a clinic or counseling office
______Business or industry
______A government agency; specify: ______________________________
______Other; specify: ______________________________

Thank You!

APPENDIX C

COMPOSITE ROTATED FACTOR MATRIX[a]

Variable	Rotated Factors								
	I	II	III	IV	V	VI	VII	VIII	IX
Reported severity of inferiority feelings in childhood (BI)	.69	—[b]	−.28	—	—	—	—	—	—
Reported amount of friction with father (BI)	.53	—	—	—	−.33	—	—	—	—
Discipline authority always one parent—rather than shared (BI)	.46	—	—	—	−.30	—	—	—	—
Reported amount of friction with mother (BI)	.46	—	—	—	—	—	—	—	—
Perceived permissiveness of parents during *S*'s adolescence (BI)	−.59	−.26	—	—	—	—	—	—	—
Reported degree of childhood happiness (BI)	−.58	—	.30	—	—	—	—	—	—
Objectivity (G-M)	−.38	−.39	—	—	—	—	—	—	—
Reported frequency of cutting classes in college (BI)	—	.51	—	—	—	—	—	—	—
Reported amount of guilt reactions to masturbation (BI)	—	.50	—	—	—	—	—	—	—
Reported amount of guilt feelings in childhood (BI)	.30	.46	—	—	—	—	—	—	—
Preference for marital counseling—as opposed to other forms (BI)	—	.42	—	—	−.38	—	—	—	—
Reported frequency of fantasy in childhood (BI)	.33	.40	—	—	—	−.32	—	—	—
Feel authors best understand human nature (BI)	.31	.39	—	—	—	—	—	—	—
MAT	—	.36	−.38	.37	—	—	—	.51	—
Birth order: eldest child (BI)	—	.32	—	—	—	—	—	—	—
Preference for leadership activities—I (BI)	—	−.42	—	−.26	—	—	−.28	—	—
Reported regularity of church attendance (BI)	—	—	.71	—	—	—	—	—	—
Religious values (A-V)	—	—	.70	—	—	—	—	—	—
Reported amount of participation in college athletics (BI)	—	—	.52	—	—	.50	−.28	—	—
Political values (A-V)	—	—	.41	—	.49	—	—	−.53	—
Religion—Protestant or Catholic vs. Jewish or none (BI)	—	—	.71	—	—	—	—	—	—
Parents' religion—Protestant or Catholic vs. Jewish, mixed, or none (BI)	−.30	—	.68	—	.37	—	—	—	—
Theoretical values (A-V)	—	—	−.53	.25	—	—	.26	.40	—
Age when sex information received (BI)	—	—	—	−.45	—	—	—	—	—
Reported amount of difficulty with math in high school (BI)	—	—	—	−.40	—	—	−.49	—	—
If can't meet deadline, tell superior (BI)	—	−.28	—	−.38	—	—	—	—	—
Economic values (A-V)	—	—	—	—	.58	—	—	—	—
Preference for leadership activities—II (BI)	—	—	—	—	.53	.50	—	—	—

Variable	Rotated Factors								
	I	II	III	IV	V	VI	VII	VIII	IX
Experimental psychologist (SVIB)	—	—	—	—	.32	—	.80	—	—
Real estate salesman ((SVIB)	—	—	—	—	.31	—	−.85	—	—
First decided to become a psychologist early vs. late in life (BI)	—	—	—	—	—	.58	—	—	—
Preference for diagnostic interviewing—as opposed to therapy or research (BI)	—	—	—	—	—	.35	—	—	—
Physicist (SVIB)	—	—	−.28	—	—	—	.91	−.31	—
Chemist (SVIB)	—	—	—	—	—	—	.90	—	—
Engineer (SVIB)	—	—	—	—	—	—	.90	—	—
Mathematician (SVIB)	—	—	−.30	—	—	—	.85	−.45	—
Life insurance salesman (SVIB)	—	—	—	—	—	—	−.91	—	—
Sales manager (SVIB)	—	—	—	—	—	—	−.80	.31	—
Guidance psychologist (SVIB)	—	−.35	—	—	—	—	−.78	.37	.36
Artist (SVIB)	—	—	—	—	—	—	.32	.83	—
Architect (SVIB)	—	—	—	—	—	—	.53	.70	—
Author-journalist (SVIB)	—	—	—	—	—	—	—	.57	—
Office manager (SVIB)	—	—	—	—	—	—	−.28	−.78	.34
Industrial psychologist (SVIB)	—	—	—	—	—	—	−.35	−.67	—
Purchasing agent (SVIB)	—	—	—	—	—	—	—	−.63	—
Social worker (SVIB)	—	—	—	—	—	—	−.25	—	.85
Public administrator (SVIB)	—	—	—	—	—	—	—	.38	.78
Social science high school teacher (SVIB)	—	—	—	—	—	—	−.41	.25	.70
YMCA secretary (SVIB)	—	—	—	—	—	—	−.40	—	.63
Personnel director (SVIB)	—	—	—	—	—	—	—	.50	.60

[a] This table is a composite of a number of rotated factor matrices; the original correlation, centroid, transformation, and rotated matrices can be found in Goldberg (1958).

[b] Only factor loadings of .25 or higher are listed.

Vol. 73, No. 13 Whole No. 483, 1959

Psychological Monographs: General and Applied

PERSONALITY, INTELLECTUAL, AND ACHIEVEMENT PATTERNS IN GIFTED CHILDREN[1]

ADMA D'HEURLE
St. Xavier College
JEANNE CUMMINS MELLINGER
Arlington, Virginia
AND ERNEST A. HAGGARD
University of Illinois, College of Medicine

ONE QUESTION that man has always asked himself is how he came to be what he is. In the present century, psychologists have approached this question from several directions; the areas of study called "learning" and "personality" are two of the most important. However, the theories and knowledge that have been developed in these areas seem inadequate when one attempts to apply them directly and with meaning to the educative process. One of our greatest needs today is to establish a set of constructs and relationships that can serve as a firm basis for educating our youth in order to utilize maximally the natural resource called human talent.

In carrying out the present research we have drawn upon the insights developed in a variety of disciplines, such as cultural anthropology, sociology, Gestalt and clinical psychology, and psychoanalysis. We drew from these sources because of our belief that, at present, no single theory or discipline provides a framework that is adequate to deal with the range of processes we set out to study. In general terms, we were interested in understanding the complex of factors associated with patterns of academic achievement in children. For example, we have endeavored to examine interrelationships among environmental pressures and controls, the developing personality structure of the children who do the learning, mental ability, and various types of academic achievement. By seeking to understand more clearly the interaction of factors relevant to the actualization of innate potential in relation to positive academic goals, we hoped to gain a clearer picture of the process of learning in the lives of the children we studied.

This report covers the first phase of a seven-year research project that was begun in 1950 in the Laboratory School at the University of Chicago. Our purpose here is to present our general procedures and findings from our study of an entire class of 76 children in Grade 3. The interrelationships among the variety of factors that we studied were traced until these children had finished Grade 9. A full report of the methodological, substantive, and theoretical findings and results of the entire study is being prepared for publication.[2]

---

[1] The writers are deeply indebted to the many persons who gave freely of their time and energy and in various ways helped in the planning and completion of this study. In particular, we are indebted to the three teachers who were most generous with their help: Suzanne Artingstall, Esther Oehring, and Barbara Stein; to Warren Syfert, then principal of the Laboratory School, and May McCord Tyler of the Records Office for their interest and cooperation; to R. D. Bock and J. J. Mellinger for their invaluable assistance in the statistical analysis of the data. We are also indebted to the United States Public Health Service and the Social Science Research Committee of the University of Chicago for making this research possible.

[2] This research was supported in part by the National Institute of Mental Health, United States Public Health Service, under research grant No. M-912, and by the Social Science Research Committee of the University of Chicago. A general report covering some aspects of this research project and an enumeration of subsequent studies is described elsewhere (Haggard, 1957). The material presented in this monograph is based on unpublished dissertations by two of the authors (Cummins, 1952; d'Heurle, 1953).

Although many previous studies have sought to relate various personality, mental, and situational factors to academic achievement, there has been little, if any, prior work reported which considers such effects and interrelationships in concert. For example, the concept of motivation has often been used in studying the relation between personality and learning, and many attempts have been made to relate strength of motivation or drive to what or how much is learned. But such attempts have been limited to the study of isolated determinants in rather restrictive situations. The present project departs from the majority of studies in this area in that it seeks to delineate the organization or structure of a variety of factors that determine patterns of academic learning.

In formulating this project, we have assumed throughout that human behavior is motivated and have sought to establish the organization of motivational or personality factors that regulate both the individual's perceptions of his world and his modes of relating to it. We have also assumed that each individual's personality structure tends to become stabilized in relation to the particular physical-social-cultural context in which he finds himself. Academic learning, then, is seen as an expression of one way the individual learns to utilize his energies—given certain innate potentials and a particular pattern of socializing pressures.

On the basis of these assumptions, we have sought answers to the following broad questions: (*a*) What effect do parental pressures have on the developing personality structure of the child? (*b*) What personality characteristics are associated with a high level of general academic achievement and, more specifically, with achievement in arithmetic, reading, and spelling in young children? (*c*) What is the relation between personality traits and general and specific measures of mental ability? (*d*) How are patterns of mental ability related to patterns of academic achievement?

Because of the complexity of these questions and because of the absence of a body of accumulated knowledge on the relationships we have sought to investigate, this research should be thought of essentially as "hypothesis generating" rather than "hypothesis testing" in a strict sense. Although in this initial study we obtained a substantial number of "statistically significant" findings, their true significance will be determined by future research in this complex but highly important area.

EXPERIMENTAL PROCEDURE

Subjects and Research Setting

Our *S*s were 76 third-grade pupils of the Laboratory School at the University of Chicago. They included all the pupils enrolled in the three sections of that grade during the school year 1950-1951, when the longitudinal study was started. These *S*s are substantially above average in general intellectual ability and academic achievement. On both measures, the median child in our sample tends to fall above the 90th percentile on national norms for children of the same age in public schools.

In socioeconomic background the children in this group also represent a relatively small segment of the total population. Almost all of them came from upper middle-class homes with a strong intellectual orientation, many being children of faculty members at the University of Chicago. The school itself is one of unusually high standards where a select staff is devoted to the education of a highly selected student body. Clearly, this is an atypical sample and, consequently, we assume that our findings will apply most directly to children from similar home backgrounds in a similar setting, and who are of comparable ability and achievement.

Actually, these children were chosen for study largely because they are typical of many of the nation's gifted children. More often than not, efforts to establish the factors related to academic achievement have concentrated on the underachiever. This emphasis is understandable in view of the pressure to solve the pedagogical problem of underachievement with all the losses it entails. Still there is a growing need to learn more about such children as our sample represents in order to better utilize human ability and to develop the intellectual resources of the society.

Our choice of the particular age level was governed by our expectations that the determinants we are interested in can be isolated with greater facility and studied with least interference in younger children, and by the fact that research on the academic achievement of children of this age has been surprisingly limited. Because of the general concern with the lack of achievement, previous work in this area has tended to center on the higher levels of education.

Grade 3 was chosen rather than any of the earlier grades because in the present system it marks the actual beginning of formal education.

By the time children reach this level they are used to school life in general. They are subjected to a higher degree of structure in school activities and a markedly greater academic orientation. The master of the basic linguistic and mathematical skills introduced in the first two grades is taken more seriously here and standardized tests in these academic areas are introduced for the first time in Grade 3.

Types of Data

Various types of data were obtained in order to provide measures of each child's academic achievement, his level of intellectual ability, the nature and extent of external pressures acting upon him, and his developing personality structure.

Academic achievement tests. To evaluate academic achievement, standardized achievement tests in arithmetic, reading, and spelling were administered to all *S*s during the first weeks of the school year (October) and again near the end of the school year (May). By obtaining two samples of achievement data we hoped to determine the stability of the achievement measure in itself, the beginnings of any systematic shifts in achievement patterns, and in their relation to the other factors investigated. The academic achievement tests given in October were: Iowa Every-Pupil Arithmetic D Elementary Form L, Chicago Reading Test B Form I, Stanford Spelling Form E. In May the following tests were used: Iowa Every-Pupil Arithmetic Test C Elementary Form M, Chicago Reading Test C Form I, Stanford Spelling Form E.

The skills tested by the above tests are: understanding of arithmetic concepts, arithmetic problem solving, reading comprehension, and the spelling of words given in context.

Mental ability tests. As part of the regular testing program in the Laboratory School, the children were given individually the revised Stanford-Binet Test, Form L. The Primary Mental Abilities (PMA) test was also administered to them as a group. The factors measured by this elementary battery are:

Verbal meaning (two tests). In the verbal *w* test, the student is required to select synonyms for given words. In the verbal *p* test, a list of words is read to the pupil, and for each word he is required to find the corresponding picture from a group of four possible choices.

Space. In this test, the student is asked to select from a group of four figures the one that would make a sample drawing into a complete square. The figures must be rotated before they will fit.

Reasoning (two tests). The reasoning *w* test requires the *S* to find the rule or principle involved in a series of words. In the reasoning test, figures instead of words are used.

Perception. In this test, the student is asked to find two drawings that are exactly alike from a series of drawings of familiar objects.

Number. This is a test of addition with numbers.

Total. The total score for each *S* is essentially the average of his standard scores on the above tests.

Parental pressures. Our attempts to obtain accurate information on the complex of parental pressures included teacher interviews with parents, questionnaires filled out by parents, interviews with the children, and such background material as was contained in the cumulative records of each child in the records office of the school.[3]

It was not possible, of course, to obtain information on many of the ways in which parents exert pressures on their children since many of these pressures tend to be communicated by subtle ways that elude the tools available to us for their appraisal. We believe that the lack of such crucial information is compensated for, in part, by information obtained in the interview with the child, the observation of his behavior, and the personality tests—since these tools give a picture of how the child perceives the adult world, the pressures that the parents exert on him, and the manner in which he responds to the various socializing forces.

Behavioral observations. Throughout the year, two of the investigators observed the children periodically in many different situations: in the classroom, during the lunch hour, on the playground, and in a variety of the less routine school activities. Our purpose was to obtain a balanced picture of the various areas of proficiency of each child and to get a full picture of how he functions in the school environment.

For the first few weeks the *O*s worked closely together in the same situations and compared observations at the end of each period. When we felt that a satisfactory degree of agreement was arrived at, this procedure was discontinued and observations were made independently. All the children were observed by each *O*, each spending a minimum of 270 hours observing the children during the school year. So much time was spent with the children because our purpose was to get to know each child well as a person; how he behaved in different situations, and what his typical reactions were to the various aspects of his school existence.

The data on parental pressures and the behavioral observations were all kept in a private file for each child. The data in this file formed the basis for judging the child on (*a*) certain personal characteristics such as spontaneity, self-confidence, persistence; (*b*) interpersonal relations with peers and adults; and (*c*) attitudes toward all of his school subjects.

A list of 27 variables covering these areas was drawn up. The variables were rated independently by two of the authors on a five-point scale in accordance with a manual prepared for this purpose.

---

[3] Copies of the parent questionnaires and the questions that were asked of each child during the interview are included in Appendix A.

When differences in the ratings were found, agreement was arrived at by further discussion. In the few cases where disagreement persisted, the average of the two ratings was used. The following observational variables were rated for each child. The *r* after each variable indicates the interjudge product-moment reliability coefficient:

1. Dependent on teacher; e.g., depends on teacher for direction and help in work; asks many questions (*r* = .95).
2. Cooperative with teacher (*r* = .88).
3. Warm and affectionate toward teacher (*r* = .88).
4. Effective as group leader (*r* = .89).
5. Effective as group participant (*r* = .78).
6. Popular with peers (*r* = .78).
7. Dependent on peers; e.g., depends on peers for direction and help in work, for affection and approval (*r* = .93).
8. Competitive (*r* = .84).
9. Hostile; e.g., gets in many fights and arguments; expresses criticism and hostile attitudes toward others (*r* = .86).
10. Spontaneous; expresses himself and acts freely (*r* = .87).
11. Active in tempo (*r* = .85).
12. Self-confident; e.g., confident in ability to do schoolwork; volunteers for difficult tasks (*r* = .85).
13. Relaxed (*r* = .80).
14. Bids for attention (*r* = .90).
15. Persistent; e.g., uses time constructively and to good purpose; stays with work until finished (*r* = .84).
16. Neat (*r* = .85).
17. Has warm and affectionate relations with parents (*r* = .81).
18. Has overprotective parents (*r* = .81).
19. Has parents who exert pressure for achievement (*r* = .86).
20. Warm and affectionate toward siblings (*r* = .88).
21. Likes arithmetic (*r* = .97).
22. Likes reading (*r* = .89).
23. Likes spelling (*r* = .93).
24. Likes art (*r* = .92).
25. Likes shops (*r* = .98).
26. Likes music (*r* = .78).
27. Likes gym (*r* = .95).

Personality tests. In the few studies that have sought to investigate the relationship between quasi-stable personality structure and academic learning, there has been a tendency to depend on a single tool for the assessment of personality factors. In the present study we have approached these factors from various vantage points with the hope of arriving at a more valid evaluation of specific personality characteristics instead of broad inferential descriptions.

The Rorschach test and the Children's Apperception Test (CAT) were given to each child to provide information concerning various important aspects of the personality structure and mode of mental functioning that the other data did not tap. These two projective tests were administered individually to all *S*s. Each test protocol was examined. Lists of variables were drawn up for each test, and their precise definition was agreed upon by the two raters.

All the records were examined, and the range of ratings was determined in the manner described. The choice of the particular variables was determined by the various aspects of the personality that we expected to relate meaningfully to academic achievement; such aspects as the child's perception of the world, of his superiors, peers, and siblings; his attitudes toward himself and others; and the areas and degree of impulse, conflict, and mastery. The following variables were rated from the Rorschach test protocols (*r*'s indicate interjudge reliabilities):[4]

1. Abstract vs. concrete approach; e.g., high proportion of *W* and lower *D;* good form quality of *W; Z,* organization of blot.
2. Orginality vs. stereotypy; e.g., high proportion of *M;* low proportion of *A;* lack of excessive *P;* variability of content categories (*r* = .88).
3. Spontaneity vs. rigidity; e.g., high *M:C* ratio; low proportion of *F* responses; high *H* and *A* in proportion to *Hd* and *Ad;* flexible sequence (*r* = .91).
4. Activity vs. passivity; e.g., type of *M;* normal proportion of *Y;* average *S* (*r* = .64).
5. Ego strength; e.g., normal proportion of *F;* lack of color and shading shock; high proportion of *H;* average *S;* high *FC* in proportion of *CF* and *C;* absence of inhibition, compulsive control, and anxiety; average distribution of locations, determinants, and content categories (*r* = .78).
6. Introversive vs. extratensive tendency *M:C* (*r* = .91).
7. Introversive vs. extratensive tendency: *FM* and *m:* Sum *Y* and *T* (*r* = .94).
8. Inhibited vs. uninhibited emotional response; e.g., low *C, M,* and *W* and high *F* ratio; excessive *F;* emphasis on black; high *A;* high *P;* high *FC* in proportion to *CF* and *C;* lack of explosive content (*r* = .85).
9. Anxiety; e.g., high proportion of *Y, V,* and *T;* high *m; x;* card description; card rejection (*r* = .71).
10. Compulsivity; e.g., high proportion of *Dd;* high proportion of *Hd* and *Ad;* large number of responses; rigid sequence; high proportion of *F* and *F* (*r* = .79).
11. Aggression; e.g., fighting and aggressive acts; implements of aggression (*r* = .94).
12. Oral concern; e.g., eating, drinking, mouths, foods (*r* = .88).
13. Genital sex concern; e.g., clothing responses; over or under reactions to sexual symbols (*r* = .71).

---

[4] The symbols used to refer to scoring categories on the Rorschach are taken from Beck (1944).

The following variables were rated from the CAT protocol (r's indicate interjudge reliabilities):

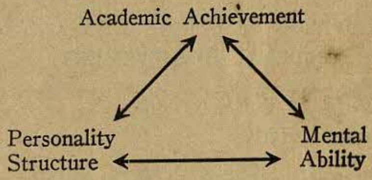
1. Activity, emotional expressiveness, and spontaneity vs. passivity and emotional inhibition; e.g., expression of strong emotions, wishes, and wants; story movement and action rather than card description; initiative and independent actions of figures identified with; emotional interaction of characters; quotes and expressions of involvement ($r = .86$).
2. Originality vs. stereotypy ($r = .86$).
3. Coherence of plot vs. disorganization ($r = .62$).
4. Complexity vs. simplicity of plot ($r = .88$).
5. Abstract concepts; e.g., generalization from stimulus detail and inclusion of several elements of picture in one concept ($r = .51$).
6. Whole approach; e.g., inclusion of major elements of picture in central story theme ($r = .57$).
7. Ego strength; e.g., recognition and solution of conflict situations; integration of emotions and actions; absence of expressions of anxiety; feelings of security and self-confidence ($r = .78$).
8. Postive vs. negative feeling tone ($r = .76$).
9. Rigidity of defenses vs. disorganization and extreme anxiety ($r = .87$).
10. Compulsivity; e.g., qualifications and indecisiveness in story; enumeration of objects, activities, etc. ($r = .78$).
11. Free-floating anxiety; e.g., expressions of unknown fears and threats ($r = .84$).
12. Helplessness of children in adult world ($r = .66$).
13. Guilt feelings; e.g., punishment themes; concern with right and wrong; concern with opinions of others ($r = .66$).
14. Rejection; e.g., fear of separation from parents ($r = .81$).
15. Aggression ($r = .85$).
16. Oral passive receptive; e.g., concern with receiving food, eating, etc. ($r = .54$).
17. Fear of oral deprivation; e.g., concern with hunger; connection of punishment and food themes ($r = .73$).
18. Anal concern; e.g., washing, cleaning, and punishment for being dirty; messing ($r = .68$).
19. Genital sex concern; e.g., primal scene anxiety; castration fear; masturbation anxiety ($r = .72$).
20. Positive vs. negative reactions to mother figure ($r = .67$).
21. Positive vs. negative reactions to father figure ($r = .79$).
22. Competitiveness with siblings ($r = .78$).

In an initial study of these variables, intercorrelations (product-moment r's) were obtained for all the variables within each list to determine possible functional duplication of variables. Six of the initial intercorrelations among the CAT scores were sufficiently high to justify combining and renaming them. This procedure reduced the original set of 28 variables to the final list given here.

The variables in each list (observational data, Rorschach, and CAT) were further correlated with the variables in the other two lists to determine the degree to which the different techniques might be measuring the same facets of the personality. This set of correlations indicated that there was little duplication and that different aspects of the personality were being revealed by the different types of data.

Analysis of Data

Before giving a step-by-step analysis of how the data were utilized, we shall restate briefly the central aim of the study and outline its basic design. Our primary objective in this phase of the research was to delineate various mental and personality correlates of the level and pattern of academic achievement. We tried to do this by determining (*a*) how mental ability is related to achievement, (*b*) what personality characteristics are associated with achievement, (*c*) to what extent personality-achievement relationships are a function of the level of mental ability, and (*d*) how do personality characteristics and mental ability relate with one another. This three-legged stool could be looked at graphically in some manner as the following:



According to this formulation the data were analyzed as follows:

1. Achievement scores in reading, arithmetic, and spelling (for the tests given in October and in May) were correlated with the personality ratings.
2. Scores on the Stanford-Binet and PMA total scores were partialled out of these correlations to determine the effect of general mental ability.
3. "Discrepancy scores" were computed by transforming the scores for each achievement test to normalized standard scores. These scores were obtained by averaging each child's standard scores and then computing the deviation of each child's standard score on each of the achievement tests from *his* average standard score. Thus, the discrepancy scores present each child's relative achievement in the three academic areas, regardless of his rank in the group. These scores were correlated with the personality ratings.
4. Achievement profiles were determined for all 76 *S*s on the basis of their scores on the October tests. The 76 *S*s were then divided by inspection into six groups according to the patterns of achievement as follows:

HA–LR High Arithmetic—Low Reading ($n = 15$)
LA–HR Low Arithmetic—High Reading ($n = 11$)

HA–LS	High Arithmetic—Low Spelling	$(n=11)$
LA–HS	Low Arithmetic—High-Spelling	$(n=13)$
HR–LS	High Reading—Low Spelling	$(n=15)$
LR–HS	Low Reading—High Spelling	$(n=11)$

Children who were assigned to the above subgroups on the basis of their achievement scores were then compared in terms of their ratings on each of the 62 personality and observational variables. Specifically, 2 × 5 contingency tables were constructed by comparing, for example, the HA–LR and the LA–HR groups on each of the personality and observational variables. Because of the relatively small number of children involved, these preliminary tables were reduced to fourfold tables by dividing each rating at its median and computing chi squares with Yates' correction for continuity.

5. Primary Mental Abilities test scores were correlated with personality and observational ratings. In order to remove the influence of general ability on these correlations, Stanford-Binet and PMA total scores were partialled out. "Discrepancy scores" were also computed from the PMA scores in the manner described above. They represent a child's relative standing on various abilities independent of his general ability score.

RELATION OF PERSONALITY AND ACADEMIC ACHIEVEMENT

Personality Traits Correlated with Achievement

Because little is known of the sets of relationships among socialization pressures, developing personality structure, and the ways in which the child learns to direct and utilize his energies toward valued goals in the world as he sees it and the degree of proficiency in the skills necessary to achieve in different academic areas, our effort has been to study these factors in concert. We have particularly investigated the relations among academic achievement and personality structure, personality structure and mental ability, and mental ability and academic achievement.

Table 1 presents results obtained when the observational and personality variables were correlated with general academic achievement and with achievement in arithmetic, reading, and spelling. Since a large number of correlations were computed, a test was made to determine whether the number of correlations reaching the .05 and .01 levels of significance exceeded the number that would be expected to occur "by chance" alone. The number of "significant" correlations not only greatly exceeded chance expectations (probability level of .001) but the obtained correlations generally make sense psychologically, and it is possible to group them for each set of variables into fairly clear-cut personality descriptions.[5]

The statements to follow pertain to characteristic trends or tendencies, i.e., *r*'s at the .05 level or higher. We present them with caution as tentative findings rather than firm conclusions. Further research with populations of different age groups and more varied socioeconomic backgrounds is required before these findings can be generalized. Together, however, the observed trends seem to add up to a total meaningful picture. The pictures that emerge for the four achievement groups differ, and presumably reflect different developing personality organizations and various approaches to academic subject matters and materials.

These results refer to the relation of personality and achievement in third-grade children only. They do not refer to the degree of stability of the observed tendencies or the amount and kind of change that may come with time.

It will be noted that Table 1 presents correlations of the two sets of achievement measures with personality characteristics. The high degree of consistency in the findings tends to emphasize the stability and the meaningfulness of the relations among personality traits and high general and differential academic achievement. Furthermore, when a difference is noted between the October and May results, it tends to be in the direction of signifying more clearly the relationship indicated by the first set of correlations.

---

[5] This test assumes that the correlations are independent. However, the observational and personality ratings and the achievement scores are not strictly independent measures so that this test is suggestive rather than strict. In the next section the same test was made for the number of correlations obtained between the observational and personality ratings and various PMA scores. In that case also, this test is suggestive since both sets of measures are intercorrelated to some degree.

High general achievers. Examination of Table 1 shows that children who receive high scores on the arithmetic, reading, and spelling tests may be characterized by traits representing exposure and responsiveness to cultural influences. Their conformity and acceptance of parental values seem to give them a high degree of security and confidence in their relation to their parents as well as to other adults. Their relationship with their teachers is also one of cooperation. These children put much effort into schoolwork and do it neatly and persistently. They are independent, self-confident, and effective participants in group activities. They also tend to be competitive with siblings and peers.

Along with the observable evidence of ego integration and emotional control are signs of inner strain. One has the impression that these children are trying hard to live up to adult expectations. They have well-developed superegos and express strong guilt feelings and concern with problems of discipline and control. Their lack of complete acceptance of, and ambivalence about, parental values appears in the unconscious expression of negative feelings toward the parents, particularly the mother.

In general, however, the general achievers show a high degree of inner harmony and are rather adept at controlling their emotions and at organizing and integrating their experiences, ideas, and feelings. The originality of their fantasy productions, capacity to handle abstract concepts effectively, and the fine balance between extratensive and introversive tendencies are characteristic of children who show a high level of over-all academic achievement.

High arithmetic achievers. Children who show superior performance in arithmetic present a personality picture similar to the general achievers, but they tend to be more spontaneous and free in their behavior. In their interpersonal relations they tend to be more aggressive, independent, and self-confident than are the children in the other subgroups. Of all organized school activities they prefer gym where they can express themselves actively and aggressively. But, in spite of their aggressive and self-assertive tendencies, they work well with others, are effective in group activities, and tend to occupy positions of leadership in the group. However, they maintain a high degree of independence in their relations with peers.

In their relations to the adult world (as well as to the peer culture) they show themselves secure and confident. In viewing their parents as authority figures and in their relations with them generally they show less strain than the high general achievers and the high achievers in reading, and greater independence than the high spelling achievers. The family picture is similar to that of the general high achievers and the achievers in reading in that the parents tend to be pressing in their demands for achievement and overprotective. Still the relation of these children to their parents tends to be more positive than that of either of the two other groups.

In their intellectual activity, the arithmetic achievers exhibit a capacity to handle details skillfully and to grasp new and meaningful relations between parts of the stimulus situation. Their mental approach is not so abstract as that of the readers, but they organize and manipulate problem materials more effectively. Evidence of the association of comfort, competence, and flexibility of organization with respect to noninterpersonal material, which Rochlin (1952) found among mathematics achievers at the college level, is observed in this group.

High reading achievers. The reading achievers turn inward and use their imagination to work out their conflicts. They are unable to express their feelings openly and are tense and strained in their behavior. They communicate little and are lacking in warmth in their relations with both adults and peers. In the peer group they are the least popular.

These children show the most intense free-floating anxiety of any group as well as the greatest fear of parental rejection and deprivation. They also are least confident in themselves. Although by no means as competitive as the arithmetic achievers, they show a high degree of competitiveness and

TABLE 1

CORRELATIONS OF PERSONALITY VARIABLES AND ACHIEVEMENT TEST SCORES FOR OCTOBER AND MAY TESTINGS

Personality Variables	General Achievement		Arithmetic		Reading		Spelling	
	Oct.[a]	May	Oct.[a]	May	Oct.[a]	May	Oct.[a]	May
A. Observations								
1. Dependent on teacher	−.23* (−.22)	−.11			−.17	−.19	−.24* (−.22)	−.07
2. Cooperative with teacher	(.23)*						.24* (.27)*	.30**
5. Effective as group participant	.20 (.21)	.24*	.19	.10	.15	(.21)	.19	.14
6. Popular with peers							.18	.10
7. Dependent on peers	−.23*	−.19	−.27* (−.22)	−.27*			−.20	−.20
8. Competitive	.26* (.23)*	.18	.42** (.38)**	.26*	.22 (.19)	.14		
9. Hostile			.22 (.19)	.13				
10. Spontaneous			.20	.10			−.10	−.23*
11. Active			.27* (.24)*	.13				
12. Self-confident	.19 (.20)	.07	.26* (.27)*	.14			.21	.08
13. Relaxed	−.19	−.24*			−.23* (−.19)	−.23*		
14. Bids for attention							−.16 (−.19)	−.19
15. Persistent	.35** (.38)**	.29*	.28* (.28)*	.31**	.23* (.23)*	.21	.39** (.39)**	.42**
16. Neat	.19 (.33)**	.09					.30** (.34)**	.36**
17. Warm and affectionate parents	−.12	−.18						
18. Overprotective parents	.18 (.18)	.20	.29* (.27)*	.22	.11	.18		
19. Parents exert strong pressure for achievement	.12	.21						
20. Warm toward siblings	−.44** (−.50)**	−.26*	−.41** (−.43)**	−.26*	−.41** (−.43)**	−.25	−.34** (−.34)**	−.34**
21. Likes arithmetic			.21	.29*				
22. Likes reading	.11 (.36)**	.29*			.14	.41		
23. Likes spelling	.25*	.14					.42** (.45)**	.38**
26. Likes music	(.21)				(.20)			
27. Likes gym			.17	.19				

TABLE 1—Continued

Personality Variables	General Achievement		Arithmetic		Reading		Spelling	
	Oct.[a]	May	Oct.[a]	May	Oct.[a]	May	Oct.[a]	May
B. Children's Apperceptive Test								
2. Originality			.09	.18				
4. Complexity of plot			.13	.20				
5. Abstract concepts	.24* (.21)	.34**	.26* (.22)	.29*	.25* (.21)	.33**		
7. Ego strength			.20	.20				
9. Rigidity of defenses	.18 (.20)	.08	.18 (.19)	.04	.19 (.19)	.06		
10. Compulsivity	(.24)*		(.23)*		.21 (.24)*	.15		
12. Helplessness in adult world	−.20	−.15			−.22	−.20		
13. Guilt feelings	.34** (.20)	.34**	.32** (.26)*	.36**	.31** (.24)*	.31**	.26* (.22)	.19
14. Rejection	−.23* (−.27)*	−.10			.19 (.25)*	.28*	−.20 (−.22)	−.27*
16. Passive oral receptive	.25* (.21)	.16	.23* (.19)	.20	.24* (.20)	.19	.32** (.25)*	.35**
18. Anal Concern	.21 (−.19)	.12	.37** (.33)**	.09	.16	.18		
19. Genital sex concern	.20	.34**	.15	.26*	.20 (.19)	.34**		
20. Positive reactions to mother	−.21	−.19	−.18 (−.22)	−.12	−.24* (−.24)*	−.24*		
C. Rorschach								
1. Abstract approach					.02	.22		
2. Originality	.19 (−.25)*	.21			.22	−.22	−.20	.08
3. Sponteneity					.19	.20		
6. Introversive M:C	.20	.02			.19	−.01		
7. Introversive FM:Y	.07	.19			−.01	.20	.07	.18
10. Compulsivity			.14	.23*			.13	.29*
11. Aggression			.19	.17				
12. Oral concern							.04	.21
13. Genital sex concern			.02	.18	.19	.22		

Note.—To save space only the variables with a correlation of .18 or higher are included.

[a] October correlations with Stanford-Binet IQ partialled out shown in parentheses.

* Significant at .05 level.

** Significant at .01 level.

ill-feeling toward their siblings. In general, they are less concerned with parental pressures and problems of control and discipline than the arithmetic group, but they exceed the other groups in their concern with sex problems.

The mental approach is strikingly abstract and theoretical, but in dealing with objective quantitative data and in organizing problem material they are markedly inferior to the arithmetic achievers. They tend to be less persistent than the arithmetic and spelling groups and show less patience with the routine school tasks. Their interest centers, rather, in the subjective symbolic subject matter areas. They prefer passive activities, such as music and reading, to all the other school activities, and they particularly dislike shop, where the manipulation and mastery of tools and materials are required.

While the arithmetic achievers may be described, in general, as outwardly oriented, this group is more inwardly bent. Although they are restricted and compulsive in their behavior, they are much more imaginative, spontaneous, original, and expansive in their fantasy life.

High spelling achievers. The spelling group tends to be passive and calm, with few evident anxieties in their behavior or fantasy. Their parents exert the least pressure for achievement, and the children tend to be more secure and passive in relation to them. Signs of ambivalence about accepting parental values are very few. They show less fear of oral deprivation, less guilt feelings, and less negative reactions to the mother and father figures than the arithmetic and reading achievers. Of all the groups, they are the least competitive with their peers and, in general, the least negative and aggressive. However, they, like the readers, express a great deal of hostility toward their siblings. The high spelling achievers relate most positively to adults and peers, but they tend to play a passive role in their social relations. In their behavior they are unobtrusive and calm and do not seek to attract attention to themselves. They are by far the most cooperative with the teacher and the most popular with peers.

Of the groups under consideration, the spelling achievers are the least active and spontaneous, and the most compulsive in their reactions. They do their schoolwork extremely persistently and neatly. They dislike art work, through which they are encouraged to express themselves freely with different media. Their fantasy productions convey the same rigidity, constrictiveness, and emotional inhibition as well as few signs of conflict and anxiety. While at the behavorial level the spelling achievers seem secure and confident, their projective records denote an immature psyche and a lack of ego strength and firmness.

The mental approach of this group is concrete and stereotyped. Whereas the arithmetic achievers show a similar tendency toward an objective, practical approach, they also manifest the power to organize parts into meaningful and complex wholes and the capacity to construct abstract concepts from concrete facts—abilities which the spelling achievers seem to lack.

The Role of Measured Intelligence

No significant change was observed in the above findings when the Stanford-Binet IQ and PMA total scores were partialled out as indicated in Table 1. This suggests that the observed relationships between personality factors and achievement are not primarily a function of measured intelligence. It will be noted, however, that a few personality factors show some relation to measured intelligence. These are: cooperation with the teacher, compulsivity, and stereotypy or lack of originality.

Role of General Achievement

In correlating the personality ratings with the discrepancy scores that represent each child's achievement in arithmetic, reading, and spelling relative to his over-all average level of achievement, we sought a cross-check on the validity of our reported findings.[6] We hoped, in particular, to determine (*a*) whether children who incline toward any one of these three areas—regardless of the level at which they achieve—also show similar personality traits, and (*b*) if so, how do they compare with the children who achieve at a high level in terms of the group norm. In other words, we wished to answer a general question such as the following: "If Johnny is an average student but does much better in one area (arithmetic, reading, or spelling), does he tend to show the same personality and behavior patterns as Ted, who is at the top of the class in the same subject-matter area?"

Since the findings from these data do not add much to the first set of correlations, we have not included them here. Their significance lies in the fact that they do not differ from (but, rather, lend support to) the results already reported. This set of correlations showed a basic similarity in the personality structure of the children who achieve in any one of the three areas (arithmetic, reading, and spelling), regardless of the level of achievement. This pattern was similar to, but not identical with, the personality picture of the children who achieved highly in the same area in terms of the group norm. The difference centered in those items that we have found to be closely associated with high achievement, such as interest in school activities, good work habits, competitive and aggressive behavior, passivity in relation to firmly directive parents, conflict over the internalization of parental values, a rigid superego, and mental activity that is marked by originality and competency in conceptual thinking.

Patterns of Achievement

For further verification of the assumption that an individual's personality structure is related to the manner and extent to which he will utilize his energies in pursuing different academic subjects, we compared the personality ratings of six groups of children who showed contrasting achievement patterns.[7] The findings obtained by this method were consistent with our other findings regarding differential achievement and tended to emphasize the differentiating characteristics of these groups.

Comparison was made between children who achieved highly in arithmetic but were low in reading (HA–LR), and those who were high achievers in reading but low in arithmetic (HR–LA). The children in the HA–LR group were found to be more responsive and spontaneous in their interpersonal relations and free in their emotional expression. They enjoyed more positive relations within as well as outside their family group, and in mastering objective experiences they showed greater skill and flexibility.

The high-arithmetic and low-spelling group (HA–LS) as compared with the low-arithmetic and high-spelling group (LA–HS) is characterized by greater spontaneity and independence. Children in the former group tend to be less passive in all their relations and particularly in their relations with their parents. They are also more academically inclined and more effective in dealing with mental tasks.

The comparison of the high-reading and low-spelling group (HR–LS) and the low-reading and high-spelling group (LR–HS) emphasizes the independence of the former, their rich fantasy life and their facility in handling verbal and symbolic stimuli.

Discussion of Achievement Results

In comparing the results of the present study with the findings from other investigations, we find the following relationships. In regard to high general achievement, Cattell (1945), Steinzor (1944), and Thompson (1948) have pointed to the high achievers' greater awareness of, responsiveness to, and conformity with environmental influences. Several other cross-sectional studies, most of which base their judgments of personality on the results of one or two objective tests, have indicated a relationship between achievement and introversive tendencies. High achievers have also been characterized by self-sufficiency, conscientiousness, persistence, a wide range of interests (particularly in the academic fields), a good adjustment in the school situation, and an urge to achieve. Other characteristics of high achievers that have been indicated are emotional tension, a high level of anxiety, good emotional and intellectual control, facility in expressing aggressive and hostile feelings, and positive relations with parents.

---

[6] The correlations of October and May discrepancy scores and personality variables are presented in Appendix B.

[7] See Appendix C.

With respect to the area of high achievers' social relations, the results of previous studies are equivocal. Gough (1949), Eckert (1935), and Neel and Matthews (1935) report a tendency on the part of achievers to be withdrawn in their social relations, whereas Steinzor (1944) and Kurtz and Swenson (1951) conclude that the achievers are better able than the underachievers to enter into personal relations and that they tend to enjoy more plentiful and supportive peer contacts. Young, Drought, and Bergstresser (1937), on the other hand, found no relation between affiliativeness and achievement.

The personality picture of the 8–9-year-old achievers obtained in the present study is, on the whole, in harmony with the trends reported here, which suggest a picture of children who have responded to cultural pressures and have made strides in internalizing adult values. The achiever's acceptance of these values is reflected in his academic interests, for example, his conscious effort to achieve in school and such related traits as compliance with authority, perseverance, and emotional control. It would seem that the achiever's acceptance of such upper middle-class values has been accomplished at the cost of a certain degree of anxiety, strain, and loss of freedom in emotional expression. However, the channeling of aggressive impulses into culturally approved behaviors and skills, and the feelings of security in relation to the adult world, which are accompanied by inner flexibility and ego integration, are signs that the high achievers have succeeded in establishing outlets that help them to cope successfully with the stresses and strains of the socialization process.

In view of the inconsistency of previous findings regarding the social interaction of the high achiever, we wish to emphasize our results in this respect. We find that, whereas the high achiever does not show warmth and emotional spontaneity in his reaction to other people, he has attained a degree of ego identity and integration which makes it possible for him to enter into effective relationships with others on a more objective footing and to be an effective group participant. This type of interaction is, in a way, as much an expression of his cultural orientation as his competitiveness. For are not the two modes of behavior, contradictory as they may seem, sanctioned by the cultural code that he apparently feels he must abide by?

With respect to achievement in arithmetic among college students, the results of the present study give support to the findings of Munroe (1946) and Rochlin (1952) concerning the objective approach of the mathematics achievers and their competence and flexibility of organization with regard to noninterpersonal activities. The fluidity of imagination, the directness of ideational reactions, the adaptability and emotional directness referred to by Cattell (1945) are also reflected in the personality picture that was found to be associated with achievement in arithmetic. Our findings indicate, further, a relatively positive relationship with the parents, independence and autonomy in behavior, effective controls, and generally good ego functions for this age level.

In comparison with third-grade spelling and reading, arithmetic requires, beyond the manipulation of objective, detailed material and the availability of past learning, a capacity to see relationships between parts of the stimulus material and to reorganize the parts into new conceptual wholes. It is hypothesized that the facility with which the arithmetic achievers express their hostile feelings in behavior makes it unnecessary for them to work out their conflicts subjectively to the degree that the reading achievers seem to do. Hence, the arithmetic achievers can work with objective factual data more freely. The security and independence that these children experience in their relations to their parents may also mean greater freedom to attempt problem solving on their own. This is quite plausible in the light of studies of the effect of institutional living on the mental development of the child. These studies have shown that children brought up in institutions that do not provide for warm and positive relating with parental figures often suffer from lack of intellectual freedom and enterprise and are lacking in their capacity to manipulate and reorganize stimulus material outside the specific context in which it is given (Goldfarb, 1943; Spitz: 1949a, 1949b).

Studies with adults have shown that *S*s with linguistic and verbal ability tend to be antisocial, inwardly bent, imaginative, and introspective. Sanford, Adkins, Miller, Cobb, et al. (1943), who also worked with children, report that those who did well in the subjects requiring verbal ability showed (besides the characteristics of the general achievers) a lack of constriction in the use of emotional material in fantasy and a predominance of antisocial fantasies. Working also with children, Levy (1933) found achievement in reading to be related to maternal overprotection, whereas Vorhaus (1946) and Stewart (1950) associate parental overprotection with retardation in this field.

In relating these findings to the results of the present study, it becomes apparent that further clarification of the relation of parental overprotection and reading achievement is needed. It will be remembered that in the present study no appreciable difference was found in the overprotectiveness of the parents of the reading and arithmetic achievers and that the slight difference that was observed weighted in the opposite direction from the one Levy indicates. The major difference we found in the child's reaction to parental control and the channels of emotional expressiveness that are available to him. We were led to believe that overprotectiveness in itself is not the main differentiating element between the arithmetic and reading achievers but, rather, the quality of the parent-child relationship and its emotional overtones. From our findings it appears that the overcontrolled child, who is shy and submissive and who responds negatively to parental control but is unable to express his feelings overtly, turns to a symbolic subjective mode of expressing his conflicts.

Reading at the third-grade level is a task that requires the substitution of one symbol for another and hence it calls primarily for response in terms of past learning. Thus, it is hypothesized that the good readers who, like the arithmetic high achievers, put great effort into their school work acquire a source of experience which explains their success in such tasks. The reading achievers, due to their constrictiveness and incapacity for overt emotional reactivity, lack the competence that the arithmetic achievers have in working with objective, factual, and quantitative stimulus materials and the active approach to intellectual data that is required in successful problem solving.

The hypothesis regarding the relation of warm emotional relating, inner integration and flexibility, and the capacity to reorganize and manipulate objective stimulus material is borne out by the findings regarding the relation of personality to reading achievement. Third-grade reading does not require the perception of relationships and the reorganization of given material that is required in problem solving in arithmetic. It does seem that, because the reading achievers lack the freedom and the inner integration of the arithmetic achievers, they are incapable of intellectual activity that requires the objective reconstruction of reality.

Language is a system of symbols into which experience is translated and articulated. This articulation of experience becomes a means of facilitating the organization, reconstruction, and re-enaction of experience and so makes it possible for the individual to master situations which, in their original form, were beyond his capacity to handle. On the basis of the personality pattern that was found to be associated with reading achievement it appears that the reading achievers turn to the verbal and symbolic areas in attempting to work out personality problems that they cannot handle more directly.

With regard to spelling achievement, it was hypothesized that because the spelling achievers are not subject to the conflicts and anxieties of the reading achievers and because they try hard to do well in their schoolwork, they achieve a high level of performance in a routine activity that requires dealing with detailed objective stimulus material and the availability of past learning. It is possible that, because the spelling achievers are passive and seemingly secure with their parents and because they are subject to few conflicts and anxieties, they, unlike the reading achievers, do not escape into the world of symbols. On the other hand, the spelling achievers lack the emotional maturity as well as the autonomy and the initiative that are associated with the active problem-solving approach of the arithmetic achievers.

Considering the difficulties involved generally in exploring the patterns of parental controls and their influence on the child's ego development, and especially the limitations of our research in this respect, we project with caution the following considerations regarding the spelling achievers. The children in this group show in their fantasy productions less concern with and anxiety about anal and sexual themes. In their relations to the parents they are considerably more passive than the children in any of the other subgroups. It would seem that their passivity and unconditional surrender to parental controls to a large measure do away with the need for preoccupation with anal and discipline themes. The passivity and lack of sexual curiosity of these children may further explain their general lack of interest in the active acquisition of knowledge. Clinical research has shown that in many individual cases extreme difficulties in learning have been traced back to the stifling of curiosity about sex. In such cases infantalism or abnormal dependency needs are often found to be parts of the general picture. Hence, it seems possible to assume that the attainment of skill in spelling is possible for the children in this group precisely because it does not require intellectual curiosity or an active approach, but rather a passive attitude and the obedient carrying out of rules learned by rote.

Summary

In brief, the results of this part of the study show that groups of children who are similar in their level and pattern of academic achievement also tend to be similar in a variety of personality traits and behavioral characteristics. Although not very pronounced, these trends form a meaningful picture which suggests that as early as the third grade, academic achievement is related to how the child perceives his world and how he relates to it, and that this relation is somewhat independent of mental ability. Although the personality structure may not fit any particular child in all of its details, it represents the traits that tend to be found in the group of children in the different achievement groups. In general, these traits are as follows:

1. The high general achievers are sensitive and responsive to socialization pressures. They have accepted adult values and strive to live up to adult expectations. Their internalization of these values seems to give them a high degree of security and confidence in their relations with adults, even though they express some underlying hostility toward authority figures. In general, however, they show a high degree of inner harmony and are rather adept at emotional control and at organizing and integrating experience. In the school situation they show a higher level of adjustment than the low achievers; they have developed good work habits and are persistent in them; they have learned to express their hostility in culturally acceptable channels and thus are able to get along well with both adults and peers.

2. The high arithmetic achievers tend to be more at peace with the world and with themselves. In their relations with their parents and other authority figures they show less strain than the high general achievers and the high reading achievers and greater independence than the high

spelling achievers. They show the greatest degree of ego integration as well as greater maturity in dealing with the outside world of people and things. They are able to express their feelings freely but at the same time they are emotionally controlled and flexible. Their intellectual processes also tend to be spontaneous and creative, and they are the most skilled in the manipulation of abstract symbols.

3. The high reading achievers seem to have withdrawn into themselves and to view the world of reality from a distance. They tend to react negatively to parental authority and are insecure in their relation with the adult world. While they find it difficult to express their feelings openly and to maintain effective relations with others, they show a great deal of freedom and expansiveness in their fantasy. Being relatively independent and free of excessive concern with adult and peer expectations, the good readers are capable of maintaining their sense of integrity by drawing on their own inner resources; with inner freedom and creativity they seek the resolution of their emotional difficulties in symbolic activity.

4. The high spelling achievers are markedly passive and dependent on adult authority. In their relations with their parents, they tend to seek approval by complete submission and strict adherence to the rules. In the school situation, they seem to be equally dependent on outside sources for direction of their thoughts and actions. They are generally less mature than the children in the other groups, both emotionally and intellectually. Their fantasy productions tend to be relatively barren, revealing a lack of inner emotional resources and intellectual creativity. The modality of their mental functioning is further characterized by inflexibility and a concrete and mechanical approach.

Considering the nature of the tasks that are mastered by the different achievement groups, these results may be viewed in the following terms:

1. An adult-like orientation and a degree of adaptive anxiety—if it is not strong enough to endanger the integrity of the ego and the individual's inner freedom—are basic factors in the academic achievement of children.

2. An active approach to intellectual tasks and the capacity to handle objective quantitative data and to reconstruct and reorganize stimulus material are associated with comfortable emotional relations with parents, freedom from free-floating anxiety, and ego integration and flexibility.

3. Inner conflict and an incapacity for emotional expressiveness tend to impede the child's freedom to deal with objective reality and his ability to reorganize stimulus material independently.

4. The passive acceptance of the forces of socialization and the lack of mature ego functions are conditions that tend to accompany good performance on routine intellectual tasks that call for a response in terms of the automatic application of past learning, but seem to interfere with a creative approach to an integration of academic materials.

RELATION OF INTELLECTUAL ABILITIES AND PERSONALITY

Personality Traits Correlated with Primary Mental Abilities

In this section we take up the relation between intellectual abilities and personality. The personality variables investigated, as well as the PMA tests used in the present study to measure intellectual abilities, are described later. The PMA tests were chosen because they measure the so-called functional unities of ability in contrast to the combination of knowledges and skills required by achievement tests.

The significant correlations between PMA scores and personality ratings are presented in Table 2. These correlations are generally smaller and fewer in number than those cited earlier between achievement scores and personality ratings. This smaller number of correlations raises the question whether this number of significant relationships might be expected merely by chance. When this possibility was tested, it was found that the number of correlations reaching the .05 and .10 levels of significance exceeded the expected number at the probability level of .001.[8] This was not true for the Rorschach variable correlations, however. Since the number of correlations between the Rorschach ratings and the PMA scores was only that expected by chance, these correlations will not be considered in our discussion.

The findings regarding the relationship between personality traits and primary mental abilities are summarized in Table 2. The results from the *verbal w* and *reasoning w* tests are so similar that they are presented together. For our sample, the correlation between these two tests was .61. Evidently when verbal and reasoning abilities are measured by word tests, at least at this age, the influence of reading ability tends to obscure any differences that may be present between these two abilities.

Following is a brief description of personality traits that differentiate the various ability groups:

PMA total. Children who score high on all primary mental abilities are outwardly independent, self-confident, and aggressive. The home picture is one of parental overprotectiveness and pressure for achievement. These children have internalized adult values and feel secure in the adult world. They are also dependent on their parents. Their fantasy stories are characterized by complex story plots and the use of abstract concepts.

Verbal w, reasoning w. These children are tense and strained in their behavior and competitive in the classroom. They express negative feelings toward their parents and siblings in their fantasies, but the negative reactions do not carry over into aggressive or hostile classroom behavior.

TABLE 2

CORRELATIONS OF PRIMARY MENTAL ABILITY TEST SCORES WITH PERSONALITY VARIABLES[a]

Personality Variables	*Vw*	*Vp*	*S*	*Rw*
A. Observations				
4. Group leader				(−.19)
7. Dependent on peers				−.18
8. Competitive	.23*			.22
13. Relaxed	−.34** (−.31)**			−.20
14. Bids for attention		.19		
16. Neat		−.20		
18. Overprotective parents	.19			.24* (.21)
19. Parents exert pressure for achievement				.21 (.21)
20. Warm toward siblings	−.25* (−.26)*			−.29** (−.30)**
21. Likes arithmetic			−.18	
22. Likes reading	.19	.37** (.27)*		
23. Likes spelling	.18 (.25)*	−.18		
24. Likes art			.24* (.26)*	
25. Likes shop	−.26* (−.25)*			
26. Likes music		(.20)		
27. Likes gym	.20	(−.25)		
B. Children's Apperceptive Test				
1. Activity, spontaneity, and emotional expressiveness			.27* (.27)*	
2. Originality			.26* (.26)*	
3. Coherence of plot	(−.20)			
4. Complexity of plot			.25* (.25)*	
5. Abstract concepts		.33** (.30)**		
6. Whole approach		.23* (.24)*		
7. Ego strength			.20 (.20)	
8. Positive feeling tone	(−.20)			
10. Compulsivity			−.21 (−.21)	
12. Helplessness in adult world	−.21	−.23* (−.19)		−.22
13. Guilt feelings	.26*		.32** (.32)**	.34** (.32)**
14. Rejection	(−.20)		.25* (.25)*	
16. Passive oral receptive		.25* (.21)		
18. Anal concern	.18			.19
20. Positive reactions to mother	−.23* (−.21)	(.23)*		
22. Competitiveness with siblings	−.20 (−.22)			−.20 (−.22)

---

[8] See Footnote 5.

TABLE 2—Continued

Personality Variables	*Rf*	*P*	*N*	Total Non-*V*	Total
A. Observations					
5. Effective as group participant	.24* (.24)*				
7. Dependent on peers	−.31** (−.25)*			−.25*	−.21
8. Competitive		.23* (.20)	.27* (.27)*		.18
9. Hostile		.28* (.26)*	.26* (.26)*	.21	.25*
10. Spontaneous		.30** (.29)**		.20	
11. Active		.21 (.19)			
12. Self-confident		.29** (.29)**	.25* (.25)*	.22	.19
14. Bids for attention		.25* (.23)*		.19	.20
18. Overprotective parents		.20	.28* (.28)*	.21	.28*
19. Parents exert pressure for achievement			.21 (.21)		.19
20. Warm toward siblings			−.19 (−.19)		−.23*
24. Likes art	.21			.22	.19
26. Likes music		.27* (.29)**	.19	.24*	.21
B. Children's Apperceptive Test					
1. Activity, spontaneity, and emotional expressiveness	.18			.21	
2. Originality	.20			:18	
4. Complexity of plot				.25*	.21
5. Abstract concepts	.23* (.21)			.20	.21
6. Whole approach		−.19 (−.20)	−.19 (−.19)		
7. Ego strength	.21	.18		.20	
10. Compulsivity	−.25* (−.25)*				
12. Helplessness in adult world	−.24* (−.22)	−.20		−.30**	−.24*
13. Guilt feelings	.21	.23* (.19)		.35**	.38**
16. Passive oral receptive		.21	.23* (.22)	.23*	.24*
17. Fear of oral deprivation		−.19 (−.23)*	−.32** (−.33)**		
18. Anal concern				.18	.20
20. Positive reactions to mother	.26* (.28)*				

Note.—Only the variables with a correlation of .18 or higher are included.

[a] See section on Types of Data for identification of symbols. Entries in parentheses represent October correlations with Stanford-Binet IQ partialled out.

* Significant at .05 level.

** Significant at .01 level.

Number, perception. These two groups show similar personality pictures. The children who score high on the number and perception tests are competitive but, unlike the high verbal group, they are outgoing, self-confident, and aggressive in their behavior. Children who score high on the perception test are further characterized in their behavior by an active tempo, spontaneity, and emotional expressiveness.

In contrast to the other groups, the number and perception groups tend to be passive and dependent and to feel secure with their overprotecting parents. In their fantasy stories both groups tend to concentrate on small details and to omit major elements of the stimulus material.

Space. In their fantasies the children in the high space group are original, emotionally expressive, and flexible; their stories are expansive and complex. Their spontaneity and emotional expressiveness seem to be inwardly directed, however, and do not appear in their classroom behavior. Unlike the other ability groups, these children tend to feel insecure and rejected in their relations with their parents.

Reasoning f. These children, like those in the high space group, are original in their fantasies, but their stories show better organization and more generalization from stimulus detail. They are self-sufficient, independent, and effective as group participants but not so outgoing and emotionally expressive as the children in the number and perception groups.

Children who score high on the reasoning *f* test show warm positive reactions toward their mothers in their fantasies. Briefly, we seem to have the picture of a child who is flexible yet controlled, both in his imaginative productions and in his personal relationships.

Verbal p. These children show immature reactions in class with a good deal of attention-getting behavior. In their fantasy stories they show a capacity to organize and integrate the stimulus material effectively.

Removing the Effects of General Ability

In order to remove the influence of general mental ability on the above relationships, partial correlations and discrepancy scores were computed. Generally, correlations obtained by these procedures seem to emphasize the previous description for each group of children. However, the following traits tended to lose their correlations with separate primary mental abilities: highly developed superego (internalization of adult values), security in adult world, and to a lesser degree, dependence on parents and overprotectiveness by parents. Apparently these traits are related to general intellectual performance rather than to specific abilities. This conclusion is further confirmed by the fact that these traits in particular showed fairly strong correlations with the total PMA score.

Discussion of PMA Results

Recently there has been an increasing interest in the relation of personality characteristics to intellectual abilities (Levy, 1933; Mayman, Schafer, & Rapaport, 1951; Munroe, 1946; Sanford et al., 1943; Stewart, 1950). It seems plausible to suppose that basic temperament characteristics may play a part in the individual's performance on intellectual tasks as well as in his nonintellectual behavior. As an example, quick reactivity to new stimuli and the ability to change set rapidly may be a basic personal trait that is shown in impulsive, spontaneous social behavior. This trait may also permit an individual to perform better on intellectual tests that call for quick perceptual reactions and rapid changes of set, whereas he does poorly on tests that require deliberate analysis before responding.

Emotional factors or personality characteristics may affect intellectual achievement in another way. Certain capacities may be developed or others neglected because of emotional conflicts within the individual. For example, a person may turn to reading as a means of relieving feelings or to escape tensions that he is unable to express or deal with directly. As a result he may perform better on verbal comprehension tests than he does on other tests.

Considering the results reported from similar viewpoints, we can suggest the following tentative explanations of the correlations between personality traits and mental abilities:

Verbal. In our sample, children with a large vocabulary tend to have negative feelings that they cannot express openly. Their behavior tends to be tense, strained, and constricted. It might be suggested that the child who is not at ease in social situations tends to withdraw from other people and to spend more of his time reading and thus performs well on a list of word meanings. Or reading may be a means by which he can release feelings that he is unable to express in behavior, and, correspondingly, his difficulty in expressing his emotions freely and may be reflected only in his tense, strained behavior. With his tension and constriction he may do better on a test that allows him to rely on previous learning and does not require new organization of the test material.

Number. This child is outwardly confident and expressive but underneath he tends to be dependent on his parents and lacking in originality. It may be that, in contrast to the highly verbal child, this child does not need to work out his conflicts in the intellectual sphere to such an extent. Therefore, he may work better with more objective, concrete data. The number test requires routine calculations and this child's relatively better performance on such a task may be related to his other manifestations of passivity and lack of imagination.

Perception. The high perception child is active, spontaneous, and impulsive in his behavior. Perhaps quick reactivity is a general trait in this child and is associated with an ability to respond quickly and accurately to stimulus detail on the perception test. Like the children in the high number group, this child expresses his emotions easily yet is dependent on his parents. These traits may be related to his facility in dealing with objective material that is familiar to him.

Reasoning f. This child is original and spontaneous in his imaginative productions.

He is also independent yet affectionate in interpersonal relations. Perhaps this child is inwardly flexible enough to do well on a task that requires operations that are new to him. In addition, he may be sufficiently independent and self-sufficient to do active problem solving and thus performs well on a test that requires reorganization of parts into new conceptual wholes.

Space. The high space child is original, spontaneous, and expansive in his imaginative productions. Again, these characteristics may be associated with the ability to perform well on the new and different task presented by the space test.

In relating the findings from this part of our study to previous research we find general consistency. A study by Pemberton (1951) related performance by adults on various perceptual and ability factors to temperament variables obtained from self-rating questionnaires. Two of the perceptual tests in her battery—space and perceptual speed—were similar to tests included in the present study. Pemberton found that *S*s who were high on perceptual speed could be characterized as impulsive, generous, unsystematic, and self-centered. These findings are somewhat similar to those obtained in this study; the perception children here were active, spontaneous, and expressive. In telling their stories the perception group were unsystematic in that they tended to omit major elements of the stimulus picture. Pemberton points out that quick reactivity seems to be a general characteristic of the *S*s who do well on space and perception tasks. This is shown by impulsive behavior and quick perceptual reactions. We might mention at this point that Gehlmann (1951), after correlating objective test scores and results from personality questionnaires, reported that *S*s who were high on perceptual factors were sociable and carefree.

The *S*s in Pemberton's space group were characterized as socially retiring, self-confident, optimistic, impulsive, and flexible. The children in the space group of the present study were found to be inwardly flexible and imaginative and not competitive in their behavior.

Other studies in this field have dealt mainly with the differentiation of verbal and mathematical abilities and their associated personality characteristics. The findings of Munroe (1946) and Cattell (1945) indicate that the person with relatively higher verbal ability tends to be imaginative, insecure, shy, and introverted, whereas the person with relatively higher quantitative abilities tends to be more objective, outgoing, and sociable. In addition, Sanford et al. (1943) report antisocial fantasies for their high verbal group. Results with the PMA verbal test support to some extent this general picture of the verbal person; the findings indicate that the high verbal child is tense and strained in his interpersonal relations and has antisocial fantasies. In the present study of third-grade children, however, the high verbal child was constricted rather than imaginative in his fantasy productions. This may be a function of such factors as the characteristics of our sample or the children's level of development.

Quantitative ability and number ability cannot be strictly equated, since the latter implies only the ability to deal with routine arithmetic calculations. It would probably be better to include reasoning ability as well, since this is partially measured by the usual mathematics test. In this study we find that the reasoning and number groups are expressive, outgoing, independent, and self-confident in their behavior. Thus, results from the present study agree with the findings of Munroe and Cattell and with our findings regarding the children who do well on arithmetic problem-solving tests.

Sanford et al. (1943), in the research mentioned above, found a relationship between achievement in art and spatial ability. It is interesting to note in this connection that the high space group of this study indicated that they like art better than did any other group of children.

In general, despite the low correlations between primary mental abilities and personality traits obtained in this study, the relationships observed generally support the results of previous research in this field.

Those writers who have listed correlations (Cattell, 1945; Pemberton, 1951) also report low correlations between personality and intellectual characteristics. It may be that personality factors play a small but consistent part in differential abilities of this type.

Summary

In summarizing our findings regarding the relationships between mental ability and personality measures, the following general statements seem to be justified:

1. Children who are inwardly imaginative and spontaneous perform better on mental tests that consist of new material and methods, tests that cannot be solved by previous learning.

2. Children who are dependent and passive, or tense and strained, perform well on tests that consist of familiar material. These children also have parents who are overprotective or exert excessive pressures on them for academic achievement.

3. Children who do not express their feelings freely do better on verbal tasks; children who are more free and active in their behavior and fantasy perform better on tests requiring the manipulation of objective data.

4. If we look at the behavior ratings only, we might form a tentative continuum ranging from rigidity and constriction to labile expressiveness. At the first extreme we would place the verbal group who are tense and constrained in their behavior. Further along this continuum would fall the reasoning group, independent and effective in classroom interaction. Finally we reach the perception group where the children are more expressive, self-confident, and spontaneous. The reasoning group falls at the midpoint, neither as tense and constricted as the verbal group, nor as active and spontaneous as the perception group.

Relation of Mental Ability and Academic Achievement

Having investigated the relations among academic achievement and personality on the one hand and personality and mental ability on the other, our next step will be to examine relations between mental ability and academic achievement.

It will be remembered that little change was observed when the Stanford-Binet and PMA measures of mental ability were partialled out of the correlations of achievement and personality (Table 1). A slight relation did appear, however, between measures of mental ability and such personality traits as confidence, independence, freedom of emotional expression, and facility in generating and using abstractions. Furthermore, in comparing the personality picture of the high ability *S*s and those who are general high achievers we find that those in the high ability group tend to be more confident, expressive, and independent in their behavior, whereas the high achievers show more tension and guilt feelings.

Considering the nature of the present intelligence tests, and in view of the persistent lack of agreement as to the nature of intelligence, it is possible to speculate about the cause-effect relation in different ways. It could be concluded, for example, that native mental capacity makes for greater confidence, freedom, and concern with right and wrong, or conversely that these personality traits make for better performance on the existing intelligence tests. It might be suggested that the children in the high ability group have not learned parental values so thoroughly as the achieving group. It may be that because they are relatively less concerned about achievement or because they do not have to work so hard to achieve academically they are less competitive, tense, conforming to adult standards and also less negative in their reactions to parents. / However, the one conclusion that can be made with confidence on the basis of these findings is that the individual's personality structure, his mental ability, and his academic functioning are so interdependent that none of them can be meaningfully considered in isolation.

Table 3, which presents the correlations of two measures of mental ability with measures of achievement in reading, arith-

TABLE 3

CORRELATION OF MENTAL ABILITY AND ACHIEVEMENT SCORES FOR OCTOBER AND MAY TESTINGS

Mental Ability	Arithmetic		Reading		Spelling	
	Oct.	May	Oct.	May	Oct.	May
Vw	.50	.43	.76	.76	.74	.65
Vp	.11	.21	.26	.42	.09	.00
V	.54	.45	.86	.78	.72	.44
S	.17	.17	−.06	−.16	.01	.07
Rw	.54	.48	.63	.61	.46	.43
Rf	.28	.40	.06	.13	.14	.17
R	.50	.52	.50	.47	.40	.40
P	.40	.24	.22	.18	.26	.09
N	.51	.42	.26	.28	.48	.43
Total NV	.50	.51	.26	.35	.34	.31
Total (PMA)	.61	.57	.47	.51	.51	.46
Binet	.34	.28	.37	.49	.20	−.06

metic, and spelling, highlights this interdependence. Here we find the various positive correlations that other observers have pointed to. These correlations are to be expected in view of the fact that intelligence tests use the same kinds of problems that are found in the standard achievement tests and have been often standardized and validated in terms of school achievement. The present correlations do not by any means rule out the influence of personality factors on either academic achievement or performance on the intelligence tests. Furthermore, if personality factors were ruled out from these correlations, and if one were to account for achievement solely in terms of the children's obtained intelligence quotient, one would expect superior performance by the children in this study in all achievement tests, as well as greater uniformity in their patterns of achievement.

A few related findings may be noted in this context regarding specific aspects of this set of correlations. It will be observed, for example, that in general the correlations of all measures of achievement with the total score on the PMA test are higher than the correlations of the same achievement scores with the Binet IQ. This is to be expected when we consider how similar the PMA tests are to standardized tests of achievement in their format, the type of questions they include, and the more extensive the use they make of acquired reading and arithmetical skills.

The difference between the two intelligence tests in their relation to proficiency in spelling is of particular interest (Binet .20 and −.06 for October and May spelling, respectively, and PMA total .51 and .46). This difference may be accounted for by the fact that the test-taking techniques, at which the good spellers at this level are expected to be more proficient, are more useful in the more "school-like" of the two intelligence tests. Achievement in reading, on the other hand, shows the greatest degree of similarity and the least amount of change with time (for October and May tests, Stanford-Binet $r = .37$ and .49, and PMS $r = .47$ and .51). It would seem that, whereas the difference in the value of test-taking technique for the Stanford-Binet and PMA tests is quite marked, the effects of proficiency in reading on the two tests is by no means so great.

As to the correlations of achievement with the individual primary mental abilities, it is not surprising to note that the good readers tend to be superior to the other two groups in the factors involving verbal skill (*Vw*, *Vp*, *V*, and *Rw*). The increased correlation between reading and the Vp factor is quite noticeable. That the ability to relate words and picture tends to improve during the eight months' interval between the two testing periods may be due to the fact that one of its important dimensions—the recognition of pictures—is a skill that children at this age have not completely mastered.

The children who achieve in arithmetic tend to show marked difference from the others in their superior performance on the PMA tests involving space perceptions and figure interpretation. The high spelling achievers tend to be more like the readers in their lack of ability to perceive and manipulate abstract spatial relationships mentally and to reason nonverbally. The relatively high correlations between spelling and number (N) is explained by the fact that the number test is a test of routine calculations, and hence utilizes similar skills to those required for good performance in spelling.

One further observation may be made regarding the good spellers' performance on the vocabulary tests (V, Vw). Although this level of performance on these tests was almost on a par with the good readers at the time of the first testing, it dropped considerably with the second testing, indicating that for them the skill of spelling may tend to become disassociated from the meaning of the words.

In general, the results of correlating academic achievement and mental ability confirms the expectation that a significant relation exists between these two sets of variables. This relation is not exhaustive and should not be interpreted to mean that intelligence, as measured by IQ tests, determines achievement test performance. This facile conclusion has unfortunately often been made on the basis of the observation of a positive correlation between performance on achievement tests and scores on the tests of mental ability—possibly because such tests as the Stanford-Binet were standardized on the basis of school grades. This view is indeed untenable in the light of our knowledge of the imperfect relation between IQ test scores and measures of achievement or performance. It is further invalidated by such findings as we have presented regarding the interrelation of both academic achievement and mental ability with personality traits.

In considering the interrelation of intelligence and performance, we suggest that thinking in terms of one active agent causing a change in another is much too simple. IQ tests usually attempt to measure something that could be called "innate potential," "facility for learning," or "intelligence"—in doing this the test makers try to eliminate reliance on previous experience or instruction. But obviously past experiences cannot be eliminated (or equated) completely. Furthermore, test-taking attitudes, temperament and personality factors, environmental and parental pressures all enter into the final test score. Thus, intelligence tests are very imperfect measures of "intelligence." Furthermore, academic achievement also is a complex function of "intelligence" and many other factors—motivation, environmental pressures and situations, test-taking facility, etc. Hence, academic achievement should not be viewed as the result of the mysterious genetic given that has been called "intelligence"; rather, both intelligence and achievement must be viewed as two complex aspects within a structure of interrelated situational and personal elements.

In summary, the results of this phase of our study may be viewed in the following ways:

1. The observed correlations between measures of mental ability and academic achievement are the result of an overlap in the two measures. Insofar as the intelligence tests measure those mental abilities that are related to academic success, and inasmuch as this measuring is done by means of acquired skills (not only linguistic and mathematical but also those pertaining to the art of test taking) it is expected that the two would show a high correlation. It is further noted that of the two tests of mental ability, the one that is more like the achievement tests in its format and its content correlates more highly with academic achievement.

2. The full meaning of the relation of mental ability and achievement appears only when we place them in the context of the total organization of experience; when we view them as related to each other and to a complex of other forces as parts of a dynamic field interacting and functioning together.

CONCLUDING REMARKS

The assumptions that academic achievement and intellectual abilities are associated with nonintellective factors, and that cognitive processes emerge within the personality organization of the individual and are meaningful in terms of his generalized effort to deal with his world as he sees it, were borne out by the findings of the present study. These results form another link in the chain of evidence that the personality is a stable configuration of dynamic processes, and is reflected and expressed in many diverse behaviors. The results of this study thus support the view that "learning," "memory," "intelligence," and "personality" are by no means isolated and independent processes, but rather they involve various expressions of the person's attempt to deal with and adapt to reality as he sees it.[9]

In relation to educational practices, the findings of this study suggest that: (*a*) personality variables should be included in any procedure designed to select and sponsor talented young persons for advanced training since the measured IQ alone does not distinguish between those who can and those who do achieve at a high level in particular academic areas; (*b*) consideration should be given to the ability and personality structures of the children who will do the learning in shaping situations to facilitate optimal learning and academic achievement; and (*c*) impediments to learning may be related to the individual's deepest motivating forces and to the early socialization pressures which help to shape his subsequent development. Ability to remove such impediments may require a thorough understanding of personality dynamics and the meaning such learning tasks have for the individual.

---

[9] As this research progresses, it may be possible to estimate the degree to which personality and mental ability measures contribute to academic achievement in this group of children.

REFERENCES

Beck, S. J. *Rorschach's test.* Vol. 1. *Basic processes.* New York: Grune & Stratton, 1944.

Cattell, R. B. Personality traits associated with abilities: II. With verbal and mathematical abilities. *J. educ. Psychol.,* 1945, **36,** 475-486.

9 x Cummins, Jeanne. Some personality and motivational factors in learning. Unpublished PhD dissertation, Univer. of Chicago, 1952.

x 10 d'Heurle, Adma. The relation of personality factors to academic learning in children. Unpublished PhD dissertation, Univer. of Chicago, 1953.

Eckert, Ruth E. Analyzing the superior college students. *Sch. Soc.,* 1935, **41,** 69-72.

Gehlmann, F. Performance on objective tests as indicators of temperament and other personality traits. Unpublished PhD dissertation, Univer. of Chicago, 1951.

Goldfarb, W. The effects of early institutional care on adolescent personality. *J. exp. Educ.,* 1943, **12,** 106-129.

Gough, H. G. Factors relating to the academic achievement of high school students. *J. educ. Psychol.,* 1949, **40,** 65-78.

Haggard, E. A. Socialization, personality, and academic achievement in gifted children. *Sch. Rev.,* 1957, **65,** 388-414.

Kurtz, J. J., & Swenson, E. J. Factors related to overachievement and underachievement in school. *Sch. Rev.,* 1951, **59,** 472-480.

Levy, D. M. Relation of overprotection and rejection to intelligence tests and school grades. *Amer. J. Orthopsychiat.,* 1933, **3,** 26-34.

Mayman, M., Schafer, R., & Rapaport, D. Interpretation of the Weschler Bellevue Intelligence Scale in personality appraisal. In H. H. Anderson & Gladys L. Anderson (Eds.), *Projective techniques.* New York: Prentice-Hall, 1951. P. 544.

Munroe, Ruth L. Rorschach findings on college students showing different constellations of subscores on the A.C.E. *J. consult. Psychol.,* 1946, **10,** 301-316.

Neel, M. O., & Matthews, C. O. Needs of superior students. *J. higher Educ.,* 1935, **6,** 29-34.

Pemberton, Carol. A study of the speed and flexibility of closure factors. Unpublished PhD dissertation, Univer. of Chicago, 1951.

Rochlin, I. Nonintellectual factors in the learning of general mathematics. Unpublished PhD dissertation, Univer. of Chicago, 1952.

Sanford, R. N., Adkins, M. M., Miller, R. B., Cobb, E. A., et al. Physique, personality and scholarship: A cooperative study of school children. *Monogr. Soc. Res. Child Develpm.,* 1943, **8,** No. 1.

Spitz, R. A. Autoerotism: Some empirical findings and hypotheses on three of its manifestations in the first year of life. In A. Freud et al. (Eds.), *Psychoanalytic study of the child.* Vol. 3-4. New York: Int. Univer. Press, 1949. (a)

Spitz, R. A. The role of ecological factors in emotional development in infancy. *Child Develpm.,* 1949, **20,** 145-156. (b)

Steinzor, B. Rorschach responses of achieving and nonachieving students of high ability. *Amer. J. Orthopsychiat.,* 1944, **14,** 494-504.

Stewart, R. S. Personality maladjustment and reading achievement. *Amer. J. Orthopsychiat.,* 1950, **20,** 410-417.

Thompson, Grace M. College grades and the group Rorschach. *J. appl. Psychol.,* 1948, **32,** 398-407.

Vorhaus, Pauline G. Non-reading as an expression of resistance. *Rorschach res. Exch.,* 1946, **10,** 60-69.

Young, K., Drought, N., & Bergstresser, J. Social and emotional adjustments of freshmen at the University of Wisconsin. *Amer. sociol. Rev.,* 1937, **2,** 166-177.

(Received September 16, 1958)

APPENDIX A

CHILD INTERVIEW

1. What do you like best about school?
2. What school subjects do you like best? (Restrict to academic.) Why? Which do you like least? Why?
3. What is the hardest subject for you? What is the easiest?
4. What is the most important thing to learn at school?
5. Which subject do you use most outside of school?
6. What things would you like to be able to do well?
7. When you have free time at school what do you like to do most?
8. What do you like to do best after school?
9. What games do you like to play best?
10. What is your favorite story?
11. What do you want to be when you grow up?
12. What do your parents want you to be?
13. If you had three wishes, what would they be?
14. How would you like to be different from how you are now?
15. If you could change and be another child, who would you like to be? Why?
16. What children do you like best in your class? Why?
17. Who is your best friend? If you had a secret, whom would you like to tell it to? Whom would you like to sit by? Play with?
18. Whom do you like least? Why?
19. What grown-up person do you like best? Who do you wish to be like when you grow up?
20. How do you like Miss X (teacher)?
21. What do you like best about her?
22. What does she do that you don't like? How would you like her to be different?
23. Do you think she likes you? Better than others? Does she like some other children better than you?
24. What do you like best about your home? What would you like to be different about it?
25. Are there some things you have to do at home that you don't like?
26. What things do you do with others at home that you like?
27. Which parent do you like to do things with best?
28. What do your parents think about your school work? How do they feel about your friends?
29. What do your parents do that makes you angry?
30. What do you get punished for? Who usually punishes you? How?
31. How much do you play with your brothers and sisters? How do you get along with them?
32. What do you and your brothers and sisters quarrel about? What usually happens?
33. Are there things that your brothers and sisters have that you would like to have? Are there things you wish you could do like them?

PARENT QUESTIONNAIRE

1. How does your child like third grade?
2. What interests him most at school?
3. What school subjects do you think he is best in?
4. What do you consider to be his outstanding abilities?
5. How do you think he feels about his abilities (as compared to the abilities of his brothers and sisters or playmates)?
6. What are his main interests outside school? How does he like to spend his free time? What activities outside school are planned for your child?
7. Which of his school subjects do you consider most important at this level?
8. What skills do you think he needs to acquire at this point?
9. What traits and attitudes are you most concerned in developing in your child?
10. What in particular do you want the Laboratory School to do for your child? How well do you think it is succeeding in this?
11. In what area do you think it has failed?
12. Please list the members of your household (include relatives, servants, etc.). Who has responsibility for your child's daily care?
13. What activities does he engage in with other members of the family? (Please check those that he enjoys most.)

Member of Family	*Activity*

14. What are some of the disciplinary problems that you have with him? How do you usually handle them?
15. What activities does he enjoy most with his brothers and sisters? How much time do they spend together? Who usually has the ideas when they play together?
16. How does he get along with his brothers and sisters?
17. What are the usual points of conflict among them? How are these settled?
18. What are the main areas of competition among them? Who is most successful?
19. What, if anything, have you done about this?
20. Please describe as fully as possible a typical day for your child.
21. Are there any additional comments that you would like to make about your child, his work at the Laboratory School, his interests, etc.?

APPENDIX B

CORRELATIONS OF DISCREPANCY SCORES AND PERSONALITY VARIABLES

Personality Variables	Arithmetic		Reading		Spelling	
	Oct.	May	Oct.	May	Oct.	May
A. Observations						
2. Cooperative with teacher					.22	.28*
6. Popular with peers			−.19		.20	
8. Competitive	.31**				−.25*	
9. Hostile	.30**				−.30**	
10. Spontaneous	.24*	.22			−.29**	−.27*
11. Active	.32**				−.29**	−.20
12. Self-confident			−.28*			
13. Relaxed			−.22		.19	
14. Bids for attention	.22				−.26*	−.26*
15. Persistent				−.20	.18	.23*
16. Neat				−.31**	.28*	.42**
17. Warm and affectionate parents					.18	.18
20. Warm toward siblings						−.19
21. Likes arithmetic	.26*	.31**			−.17	−.29**
22. Likes reading		−.18		.50**		−.31**
23. Likes spelling		−.18		−.24*	.35**	.42**
25. Likes shop	.30**	.29**	−.25*			
27. Likes gym						−.18
B. Children's Apperceptive Test						
5. Abstract concepts			.19		−.20	−.23*
6. Whole approach	−.18		.38**	.24*		
7. Ego strength		.20			−.18	
14. Rejection						−.18
17. Fear of oral deprivation			.24*			
18. Anal concern	.21				−.30**	
19. Genital sex concern				.22		−.25*
C. Rorschach						
1. Abstract approach				.32**		−.23*
2. Originality	−.19					
3. Spontaneity			.27*	.24*		−.18
4. Activity			.24*	.20		
6. Introversive M:C		−.18				
10. Compulsivity			−.18	−.33**		
11. Aggression	.19				−.28*	−.18

Note.—Only the variables with a correlation of .18 or higher are included.
* Significant at .05 level.
** Significant at .01 level.

APPENDIX C

COMPARISON OF GROUPS SHOWING DIFFERENT PATTERNS OF ACHIEVEMENT

Personality Variable	Level of Significance of Chi Squares					
	HA-LR v. LA-HR		HA-LS v. LA-HS		HR-LS v. LR-HS	
	Oct.	May	Oct.	May	Oct.	May
A. Observations						
1. Dependent on teacher						−.10
2. Cooperative with teacher		.07	−.10	−.08	−.10	−.08
3. Warm toward teacher	.10				−.10	
4. Group leader	.10	.10				.05
6. Popular with peers	.20	.12	−.10		−.07	
7. Dependent on peers		−.03	−.05	−.07	−.10	
8. Competitive					.10	
9. Hostile			.03	.09		
10. Spontaneous	.20		.02			.03
11. Active	.08		.04			
12. Self-confident		.10				.10
13. Relaxed				.05		
14. Bids for attention						.07
15. Persistent	.05	.03			−.08	
16. Neat	.10	.15	−.06	−.05	−.06	−.01
19. Parents exert pressure for achievement			.02			
20. Warm toward siblings			.20	.15	−.10	−.05
21. Likes arithmetic			.07	.05	−.05	
22. Likes reading				.05		.01
23. Likes spelling			−.10	−.10	−.05	−.01
24. Likes art					.05	.10
25. Likes shop		.07				
27. Likes gym			.10			
B. Children's Apperceptive Test						
3. Coherence of plot				.15		.10
5. Abstract concepts			.10	.05	.05	.03
6. Whole approach	−.10	−.10	.06	.02	.07	.03
10. Compulsivity	−.10	−.10				
11. Free-floating anxiety					.05	.05
12. Helplessness in adult world	−.05	−.08			−.10	−.06
13. Guilt feelings				.10	.10	
14. Rejection			.10	.10		.10
15. Aggression	.05		.05			
17. Fear of oral deprivation		−.05				
19. Genital sex concern	−.10					
20. Positive reactions to mother		.18				
22. Competitiveness with siblings	−.10	−.15				
24. Grammatical level			.07	.05	.05	.05
C. Rorschach						
1. Abstract approach				−.08		.05
3. Spontaneity						.05
4. Activity	.10			.10		
6. Introversion M:C	−.05	−.13			.08	.05
8. Inhibited emotional response		.10				
10. Compulsivity		−.10		−.15	−.10	
11. Aggression			.01	.05		
13. Genital sex concern	−.10					

Note.—To save space only *p* values of .20 or less are given.

Vol. 73, No. 14 Whole No. 484, 1959

Psychological Monographs: General and Applied

IMPAIRED PERFORMANCE WITH ACUTE SLEEP LOSS[1]

HAROLD L. WILLIAMS, ARDIE LUBIN, AND JACQUELINE J. GOODNOW

Walter Reed Army Institute of Research

THE INITIAL PROBLEM in studying sleep loss is the difficulty of finding any performance, either psychological or physiological, which is consistently impaired. More than 60 years after the first experiment in 1896 (Patrick & Gilbert), a recent review of performance on motor and mental tests concluded that "subjective attitude" (appearance, mood, and behavior) is the only factor seriously affected by sleep loss (Tufts College, Institute for Applied Experimental Psychology, 1949, p. 1). For physiological performance, most of the functions tested show a similar lack of impairment (Kleitman, 1939; Lee & Kleitman, 1923; Tyler, Goodman, & Rothman, 1947) and the one measure which has shown a consistent effect—changes in alpha rhythm (Bjerner, 1949; Malmo, 1958; Tyler et al., 1947)—presents a number of apparent inconsistencies in its relationship to psychological performance.

In response to these difficulties, the most frequent proposal is that the appropriate task or function has not yet been found. For psychological performance, for example, it has been argued that tests should be nonmotor (Weiskotten & Ferguson, 1930) more complex (Laslett, 1924), more demanding of higher mental activity (Kleitman, 1939), or, at the other end of the scale, less challenging (Wilkinson, 1957), less motivating (Ax, Fordyce, Loovas, Meredith, Pirojnikoff, Shmavonion, & Wendahl, 1957), or longer and less amenable to compensatory effort (Edwards, 1941). This search for a sensitive task was not rewarding either in past studies or in our own first efforts. We have come to adopt instead an approach found in studies of fatigue (Bartley, 1957; Bartley & Chute, 1947; Welford, 1953). With this approach, the most general problem is to identify and measure the sensitive aspect of performance on any task. The more specific problems, together with the sources from which we have borrowed some solutions, may be summarized as follows:

1. What is the over-all nature of impairment? The recent emphasis in studies of fatigue has been on the increasing unevenness of performance rather than on any gradual and continuous decline. There are several ways of viewing unevenness. We have followed Bills (1931) in regarding it as the result of blocks or brief periods of no response which increase in frequency and duration with continued mental work, performance between blocks being maintained at the initial level or close to it.

The idea of impairment as brief periods of no response has been extended to sleep loss on two previous occasions, both times with profit. Warren and Clark (1937) used Bills' definition of a block as a pause in responses that was more than twice as long as the average response time. They found that the number of blocks showed a decided increase with sleep loss, while the customary measures of modal or average response gave the usual result of little or no effect. The concept of blocks was taken up again in an excellent but little-known study by Bjerner (1949). Bjerner used the somewhat more general term "lapses" instead of "blocks," and found that lapses in the form of long reaction times were significantly associated with changes in alpha rhythm and pulse rate. Largely on this basis, Bjerner regarded lapses during sleep loss as reflecting brief periods of sleep or a condition like sleep, an approach which we have followed in the present study.

---

[1] This monograph is part of a two-year study of the effects of sleep loss on performance, personality, and social interaction. Our debts are numerous. To single out only some of them, we are indebted to Charles Gieseking, Irvin Rubinstein, and Betty Shanks for a major share of test administration; to Ethel Eldridge and Ometta Kearney for both test administration and data analysis; to Herman Becker and Pearl W. Mack, who, together with members of the hospital nursing staff, were invaluable in their constant care of *S*s; and to David McK. Rioch, George Crampton, John Armington, and Edward Murray for constructive comment and practical assistance.

2. In what aspect of performance will impairment appear? In the studies so far cited, lapses were related to changes in speed and only indirectly to the occurrence of errors. In the work of Broadbent (1953), however, both types of change were tied to the way in which lapses affect performance on different types of task. As a rule, tasks are grouped in terms of the common mental process which they are thought to represent—all learning tasks, for example. In contrast, Broadbent's grouping is in terms of whether tasks are "paced" or "unpaced." Unpaced tasks are those like reaction time, where the stimulus remains present until the *S* makes a response, or like most problem-solving tasks where the *S* can determine when he will proceed from one part of the problem to another. In all of these tasks, a response can be deferred until a lapse is over, so that impairment will appear as a change in speed or output but not in accuracy. On paced tasks, however, the stimulus is present for only a short time and the response must be made within a brief and prescribed interval: a vigilance or a radar task, for example. Under these conditions, a lapse coinciding with the appearance of the stimulus can produce an error, particularly an error of omission.

With one modification, we have adopted Broadbent's division of tasks. We have preferred to group tasks in terms of their being *S*-controlled or *E*-controlled, with the pacing of responses as one critical feature—certainly a major one—that may be controlled either by *S* or by *E*.

3. What conditions promote or modify the appearance of impairment? The negative and inconsistent effects of sleep loss have often been attributed to variations in such conditions as task duration, stimulus monotony, knowledge of results, and the challenge of a task. It is important to separate these conditions from the problem of measuring the sensitive aspect of performance because failure to measure the sensitive aspect can, under any condition, result in a spurious picture of no impairment. In the two kinds of task, *S*-controlled and *E*-controlled, we studied the effects of two of these conditions: (*a*) knowledge of results, and (*b*) uninterrupted and unchanging stimulus and response conditions.

The remainder of this monograph is devoted to exploring these general ideas and to determining how well they account for the apparently inconsistent effects of sleep loss. We hope to state them fully enough so that they can be used as a guide in testing for impairment and as a starting point for relating changes in test performance to physiological changes, notably to changes in the EEG alpha rhythm.

PROCEDURE AND DESIGN

The results are drawn from two studies of sleep loss, one in 1956 and the other in 1957. There are some differences between these two in the sample of *S*s and in the design. In both studies, however, the experimental setting and general procedure reflect our aim of providing a noncoercive, nonrestrictive atmosphere with a minimum of heavy physical work.

Design. Both studies were designed so that the *S*'s performance could always be evaluated against two standards: (*a*) the *S*'s own performance before sleep loss and after recovery, and (*b*) the performance of control *S*s who were given the same tests but allowed to sleep 8 hr. each day.

The general plan for both studies covered an 11-day cycle and is given in Table 1. The major design differences between 1956 and 1957 were two. In 1956, the sleep loss goal was 72 hr.; in 1957, 98 hr. In 1956, all *S*s were run through two 11-day cycles; the experimental *S*s in the first cycle became controls in the second and vice versa. This was not satisfactory, as experimental *S*s tended to lose interest after what they deemed their real job was completed. Accordingly, in 1957, *S*s went through only one cycle and the next cycle started with a fresh batch of *S*s.

At the beginning of a cycle, a group of 10 to 12 men was divided into experimentals and controls on the basis of one or two days of screening tests. In 1956 the division was based on a sociometric variable. In 1957, two *S*s who met an EEG

TABLE 1

General Design for Sleep Deprivation Experiments

Group	Period		
	Baseline	Sleep Loss	Recovery
	Days 1–4	Days 5–8	Days 9–11
Experimental	Normal wake-sleep pattern	74-98 hr. sleep loss	Normal wake-sleep pattern
Control	Normal wake-sleep pattern	Normal wake-sleep pattern	Normal wake-sleep pattern

criterion were selected as experimentals. The remaining *S*s were matched as far as possible on the basis of scores on the Army Classification Battery (ACB) and then assigned randomly to experimental or control groups. The 1956 *S*s were not told into which group they fell until the end of the baseline period. They could, however, easily infer this from the location of their sleeping quarters and in 1957 we made this information more formal by posting the division of *S*s on the second baseline day.

Procedure. To avoid diurnal variations, any particular test was given at the same time of day for each testing, and the time of testing for experiment and control *S*s was distributed as evenly as possible. The major exception was one occasion in the 1956 study when the experimental *S*s were tested in the early hours of the morning after approximately 70 hr. of sleep loss.

Because of time limitations for both *E* and *S*, not all *S*s were given all tests. With the exception of the selection of *S*s for EEG testing, *S*s were assigned to tests on the basis of availability and, to a lesser extent, ACB scores.

Subjects. In both 1956 and 1957, the *S*s were volunteers from groups of army enlisted men. In 1956, 24 *S*s came from a group of religious conscientious objectors. In 1957, 50 *S*s came from a regular field hospital unit stationed at Fort Belvoir, Virginia. The age range for the 1956 group was 19–22 yr.; for the 1957 group, 18–45 yr. Since one might expect differences between the performance of these two samples, all results presented will be labeled according to the year in which they were gathered.

Experimental setting. The experimental ward was a self-contained unit that consisted of dayrooms, a nurse's office, and sleeping quarters for the experimental *S*s. Control *S*s spent their days on this ward, but slept in a separate ward. All testing was conducted on two wards adjoining the experimental one.

The dayrooms were equipped with television, radio, record players, table tennis, and a variety of craft and hobby kits. Extra food was always available. In 1957 coffee was also available. (Because of their religious views, the 1956 *S*s did not drink coffee.)

A 24-hr. watch was maintained to prevent *S*s from sleeping or napping. Part of this watch was kept by a staff of three nurses and six corpsmen who worked on eight-hour shifts. In addition to these, project administrative officers and control *S*s were often available. The emphasis throughout was on persuasion. If an *S* dozed, he was awakened immediately, and, outside of testing sessions, was taken for a walk, engaged in games, given a shower, etc. As it turned out, much of this work was taken over by the *S*s helping each other and trying to avoid the intervention of "assistants" who were not from their own unit. On the ward, this was made easier by the concentration of *S*s in a limited number of rooms. In the testing rooms, and on the way to and from these, either *E* or an assistant was continually present.

Motivation. The following details are given in view of the emphasis on motivation as a major determinant of whether or not impairment appears (Ax et al., 1957; Wilkinson, 1957).

The *S*s received, in addition to their regular base pay, the relatively small amount of twenty-five dollars. Most of their interest, however, came from competing with one another and from turning the ability to endure sleeplessness into a question of "guts." This picture of a general high level of motivation is supported by three points:

1. All *S*s completed the task; and in 1956, a number of *S*s requested that they be allowed to see how long they could hold out, a request we acceded to by allowing some of them to continue for 89 hr. (the initial goal was 72 hr.). This is in contrast to Tyler's (1946) study where 15% of the *S*s dropped out before the end of 56 hours.

2. The *S*s denied their sleepiness. On scales for both sleepiness and fatigue, the 1957 *S*s rated themselves as more fatigued than sleepy, and gave higher ratings of sleepiness to their fellows than to themselves. The self-ratings of sleepiness were

considerably below the ratings given by observers (Murray, Williams, & Lubin, 1958).

3. Incidents of anger or irritability were, to our surprise, rare. Instead, the *S*s' moods were most often marked by listlessness and a sense of simply putting one foot after the other in a dogged attempt to keep going. These results are similar to those reported by Eagles, Halliday and Redfearn (1953) for a pair of cooperative and gently handled *S*s.

Methods of analysis. Whenever possible, the average Spearman rank-order correlation with hours of sleep loss was computed. The a priori rank order was determined by using the hypothesis that impairment would increase as hours of sleep loss increased. The distribution of this rank-order coefficient, which we have called K, is described by Lyerly (1952). Occasionally, to measure the size of change regardless of trend, the measure K was supplemented by pooling baseline and recovery results and comparing them to the sleep-loss period. An average Kendall tau rank-order coefficient was computed. The distribution of this coefficient, which we call T, is described by Jonckheere (1954). For all results, an asterisk after a value of K or T refers to values at or beyond the .05 level of confidence.

Hours of sleep loss or wakefulness were calculated from 0600 on the day which ended in the *S*'s first sleepless night. For recovery testing, the reference R_1 throughout refers to the day after the *S*'s first sleep. As is typical in experiments on sleep loss, this sleep was little different in length from the *S*'s usual number of hours of sleep.

SUBJECT-PACED TASKS[2]

In a task which is completely *S*-controlled, the *S* determines the time when the stimulus appears, the duration of the stimulus, the time within which the response must be made, the distribution of work and the acceptable level of performance.

In the present section we shall be mostly concerned with tasks in which the effective response time and, indirectly, the duration of the stimulus are *S*-controlled. As a general term, we shall refer to these tasks as *S*-paced. On all of them, we expect to find changes in speed.

One of these *S*-paced tasks is simple reaction time where the lapse hypothesis allows us to predict a particular form for the changes in reaction time; namely changes in the range and distribution of response times. Three of the remaining tasks, like addition or concept attainment, illustrate an impairment in speed but not in accuracy. And the final task points to changes in the quality of performance that occur when the *S* can control the standards for performance.

Reaction-Time Tasks

Typically, a reaction-time test consists of a warning signal and a foreperiod followed by a signal to which the *S* responds as rapidly as possible. The *E* controls the time when the warning appears, the length of the foreperiod, and the time when the signal appears. The duration of the signal, however, and the effective response time are *S*-controlled. The signal generally remains present until the *S* responds. The time allowed for response is either infinite or, if controlled, so generous that it extends well beyond the time *S* takes for any response. On a task like this, the usefulness of the lapse hypothesis lies in predicting the form taken by changes in reaction time.

The Nature of Decrement

Suppose that lapses with sleep loss show the same increase in frequency and duration that Bills (1931) reports for prolonged, uninterrupted work. When a lapse coincides with the appearance of the signal, the response will be delayed. When the signal occurs between lapses, the speed of response will be unaffected. There will be then some trials on which reactions will be as fast as any during the baseline period, and some trials on which reactions will be considerably slowed. The difference between the slowest and fastest responses will become progressively greater as lapses increase in duration. As lapses increase in frequency, the distribution of reaction times should become positively skewed. Gradual impairment (the obvious alternative to the lapse hypothesis) implies a constant increase in all reaction times, with no change in the range and distribution of responses. Both hypotheses imply an increase in mean reaction time.

---

[2] For the administration and analysis of these tests, we are indebted to Ometta Kearney and Ethel Eldridge (reaction-time tasks), Betty Shanks (concept attainment), and Irvin Rubinstein (predicting probable events).

In all four of the reaction-time tasks we used, the increase in range and in positive skew is the most striking result. There is as well a slight over-all increase in reaction time so that the fastest responses during sleep loss are slightly slower than the fastest in the baseline period.

To illustrate these effects, the results of a reaction-time task, given in 1956 to eight experimental *S*s, are presented below.

Apparatus and procedure. The *S* sat at the center of a table. Before him was a center light (blacked out), two response keys (left and right), and a small loudspeaker. To the extreme left and right of *S* were two lights (each 23 in. from the center of the table). *S* fixated the center spot. A click from the loudspeaker warned *S* that in 2 sec. one of the two sidelights would be turned on. With the appearance of a light, *S*'s task was to press the corresponding key (left or right) as rapidly as possible. Pressing the appropriate key turned off the light. The length of the warning period (2 sec.) was controlled by a Microflex Timer. Reaction time was measured by a Standard Precision Electric Timer (SPET).

Both the Timer and the SPET were operated by *E* from an adjoining room with a thick wall between *E* and *S*. The occasions of testing are shown on Fig. 1. Each testing session covered 72 trials.

Results. Figure 1 shows a marked increase in reaction time during sleep loss, whether we consider the mean or the median. After 78 hr. of sleep loss, mean reaction time is twice as long as on the last baseline day. Statistically, reaction times during the sleep-loss period are significantly longer than during the baseline and recovery period (tau [T] for mean reaction time is .79 for the mean and .60 for the median). Also, increases in reaction time are strongly related to increasing hours of sleep loss (K for the mean was .82, for the median, .85).

Of most interest to us at the present moment is the way in which various trials contribute to these general measures (mean or median). In Fig. 2 we present the average of the 10 shortest and 10 longest reaction times through the baseline, deprivation, and recovery periods. Figure 2 leaves us in no doubt that *S*s during sleep loss are capable on some trials of coming very close to their best performance during the baseline period. The 10 best trials during sleep loss show only a slight increase over those during the baseline period. The increase, however, is consistent (T = .65, K = .85). Far more striking is the enormous increase in the duration of the long reaction times. The *S*'s poorest performances become progressively worse, even though his best performances remain close to his original level.

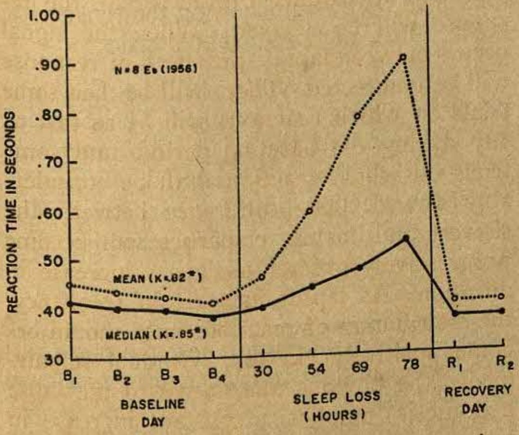


FIG. 1. The effect of sleep loss on a reaction-time task. [United States Army photograph]

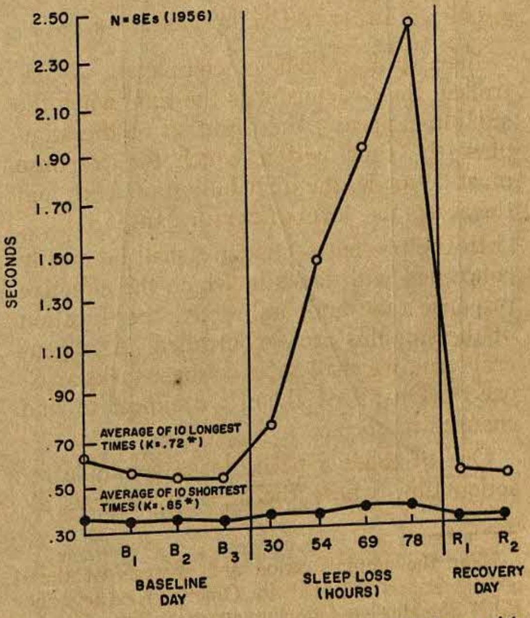


FIG. 2. Changes in extreme reaction-time with sleep loss. [United States Army photograph]

This increase in range of reaction times can be regarded as support for the argument that with sleep loss lapses increase in duration. Do they also increase in frequency, as Bills (1931) found for fatigue? If we apply Bills' definition, a "lapse" or "block" would be any reaction time that is twice as long as *S*'s average reaction time on the last baseline day. Figure 3 shows that these lapses occur more and more often as sleep loss increases.

These increases in the duration and the frequency of lapses are found consistently on all four of the reaction-time tasks that we used.

Task Conditions Affecting Decrement

Task duration. The four reaction-time tasks we used varied in several ways: the number of trials given, the regularity of the foreperiod, and the duration of the foreperiod. When the experiment was designed, our concern was predominantly with the latter two conditions.[3] As it turned out, the reaction-time data, when analyzed with respect to the regularity and duration of the foreperiod, gave inconsistent results. In contrast, we consistently found that reaction time was an increasing monotonic function of task duration (a variable which was not directly considered in the experimental design).

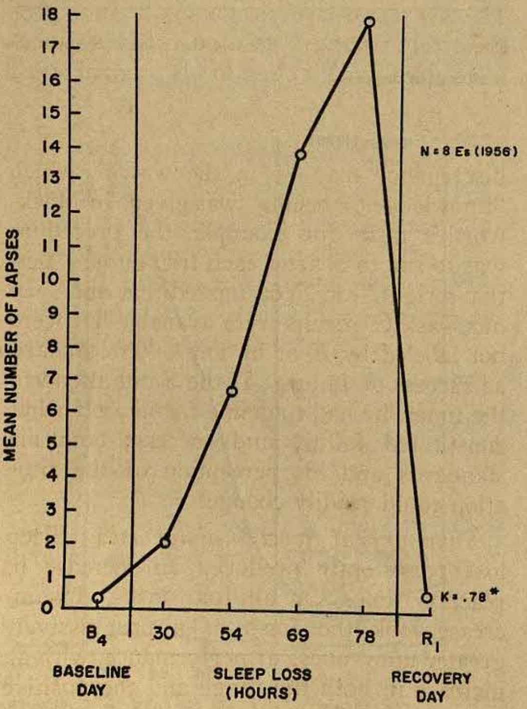


FIG. 3. Incidence of long reaction-times (twice the baseline average). [United States Army photograph]

Table 2 presents a summary of results and conditions in the four tasks. Task A, which ran for 72 trials without interruption, was highly sensitive as a whole to sleep loss. In the other four tasks (B, C, D, and E), there was a 1-min. break between conditions and some changes in conditions after the break. Here there was a tendency for only the last part of the task to be sensitive, regardless of the variations in other conditions.

The critical factor appears to be the length of time, without interruption, that the same stimuli occur and the same response is required. Breaks in a task appear to delay the occurrence of lapses until the later stages of the task. From our data it is not clear, however, whether the time break derives its effectiveness from being simply a rest period or from introducing a welcome change into the stimulus situation. The greater importance of the latter is certainly implied by the emphasis on stimulus and task monotony in a number of studies of fatigue (Broadbent, 1953a; Mackworth, 1950; Solomon, 1948), and it fits with Wilkinson's (1958) emphasis on the sensitivity, to sleep loss, of "predictable" tasks.[4]

Knowledge of results. In Task C, *S* was given knowledge of results. The *S* faced SPET. For half the trials, this timer was in circuit so that it registered, within *S*'s view, the reaction time for each trial. The *S* reset the timer after each response. For

---

[3] At the time, we were following previous work (Rodnick & Shakow, 1940) that had emphasized the relevance of these variables to the *S*'s ability to maintain a set.

[4] More detailed material on the effects of task interruptions is presented in the section on *E*-paced tasks.

TABLE 2

DESCRIPTION OF REACTION-TIME TASKS

Task	*S*s	Stimulus (St) and Response (R)	Order and Nature of Conditions				Results	
			Foreperiod	Knowledge of Results	Trials for Each Condition	Approx. Task Time	Tau	K
A	1956 Es–8	St—1 of 2 lights R—Press 1 of 2 keys	Constant (2 sec.)	No	72	10′	.79[e]	.82[e]
B	1956 Es–10	St—1 light R—Press key	1. Irregular[a] 2. Regular[b]	No No	50 50	30′ 30′	.52[e] *.63*[e]	.40[d] .95[e]
C	1957 Es–8 Cs–8	St—1 light R—Release key	1. Regular 2. Irregular 3. Regular 4. Irregular	No No Yes[c] Yes	25 25 25 25	30′ 30′ 30′ 30′	.23 .20 .12 *.62*[e]	.25 .12 .62[e] .81[e]
D	1957 Es–4 Cs–4	Same as C	1. Regular 2. Irregular	No No	50 50	30′ 30′	.50[d] *.80*[e]	.50 .50
E	1957 Es–6 Cs–6	Same as C	1. Irregular 2. Regular	No No	50 50	30′ 30′	.53[e] *.84*[e]	.75[e] .83[e]

[a] Intervals were either 5″, 1″, 15″, 2″, or 25″, in random order.
[b] Interval constant within each block of 10 trials; for successive blocks, 5″, 1″, 15″, 2″, 25″. *S* is told that after each block the interval will change.
[c] Timer registers *S*'s reaction time within *S*'s view. *S* resets timer after each response.
[d] Significant at .05 level.
[e] Significant at .01 level.

the other half of the trials, the clock was not in circuit. The instructions about the use of the timer were simple: "Whenever you finish lifting your hand off the key, look at this clock on the left. That tells you your reaction time. Fifty means you took one-half second; 30 means you took three-tenths of a second, and so on. Before putting your hand back on the key, reset the clock." These instructions leave a great deal up to *S* and some of the *S*s regarded resetting the clock as a purely motor part of the task and a distraction to boot. In this particular form, knowledge of results did not help the *S* to resist the effects of sleepiness (see Table 1).

This result was surprising, in view of the emphasis placed on knowledge of results by, for example, Mackworth (1950) in his fatigue studies and by Wilkinson (1957) in his studies of sleep loss. The source of this discrepancy may lie in the way in which "knowledge of results" was given. In Mackworth's task, for example, the procedure was to say to *S* after each trial either "Yes, that's right," or "You missed that one." In our Task C, results were available but were not labeled by *E* or by any other standard as success or failure. If the *S* did attend to the timer, he had to define for himself what constituted failure and success; both his standards and his perception of the situation could readily change.

Summary of reaction-time tasks. Sleep loss persistently produced an increase in reaction time. On all four tasks, this increase took the form of a progressively greater unevenness of performance, with an increase in both the range and the positive skew in the distribution of reaction times. These results follow from and agree with the hypothesis that sleep loss produces an increase in the frequency and duration of lapses. In contrast, there may be a gradual deterioration affecting every reaction time, but it is very slight.

The major condition promoting impairment is task duration, particularly the length of time that a task continues without interruption, and without a change in either the stimulus conditions or in the response required.

Tasks in Which Speed and Accuracy Can Be Measured Simultaneously

In a simple reaction-time task, speed is usually the only appropriate measure of performance. In more complex tasks where *S*, for example, has to discriminate between different stimuli, has to combine information, or make inferences, there are usually several ways of measuring efficiency. Since these measures may be relatively independent of one another, it is possible for one aspect of performance to decline while another remains stable, or even improves. The two aspects usually measured are speed and accuracy.

How does it follow from the lapse hypothesis that, in *S*-paced tasks, accuracy will be less sensitive to sleep loss than speed? Three assumptions are involved: (*a*) Between lapses, performance can be maintained at close to the normal level. At this level, lapses simply represent "time-out" between responses. (*b*) If a lapse interrupts a chain of operations so that the *S* loses his place or makes an error, the *S* can use his unlimited response time to start again or to correct the error. (*c*) In most tasks where both speed and accuracy are measured, the task-set is such that the *S* will sacrifice speed for accuracy.

To determine the usefulness of these assumptions and to check the general expectation of an impairment in speed rather than accuracy, we present below the results of three *S*-paced tasks: adding, concept attainment, and sending instructions.

An adding task. The main instructions were: "Here is a list of numbers (one-digit). I want you to add each two numbers and write the answer here. . . . You will be doing this for three minutes. . . . Just keep going, as fast, but as accurately as you can." The test was given both in the morning and evening, and Fig. 4 shows the results for evening testing. Even after 86 hr.

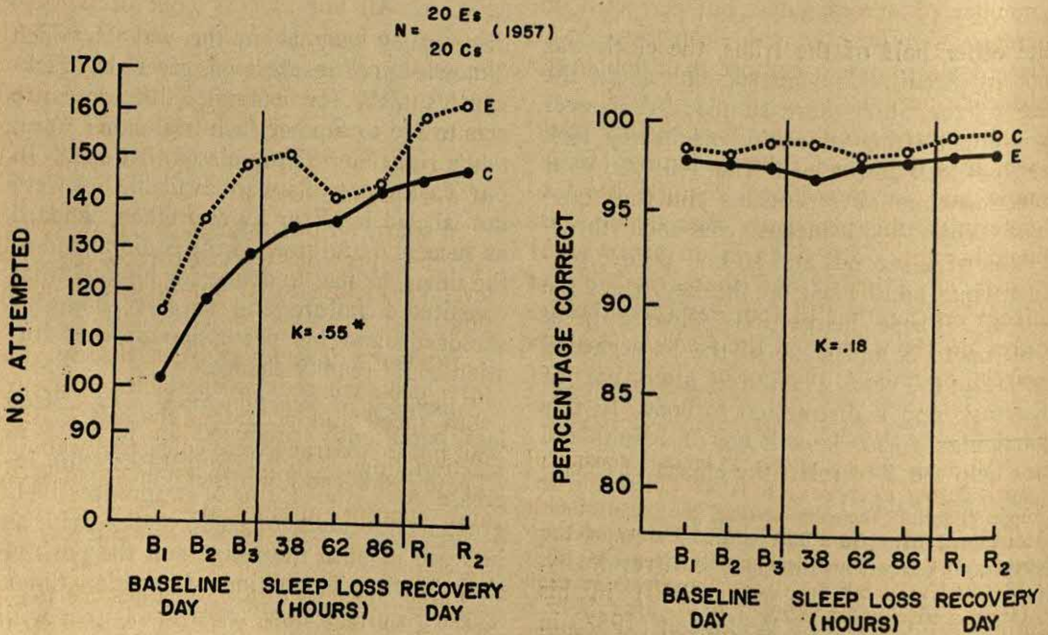


FIG. 4. Performance on an adding task. [United States Army photograph]

of sleep loss, the percentage correct shows minimal change. The number of pairs attempted, however, decreases significantly after 48 hr. of sleep loss.[5]

A communication task. In this task, designed and already reported by Schein (1957), the *S* sent and received instructions about the placement of domino-like pieces of cardboard into a pattern (10 or 25 pieces to a pattern). The sending part of the task was essentially *S*-paced. The *S* sat before a pattern of pieces. His task was to tape-record instructions that would allow someone else to build the pattern. The *S*s had already become practiced in a "code" for giving these instructions, describing each piece as placed either horizontally or vertically, and specifying its point of contact with some previously placed piece. The *S*s were told that they could work as fast as they wanted to but that they would be scored on accuracy; if they made an error or omitted something, they should correct it or fill in the necessary information.

In this task, the number of pieces described incorrectly shows an insignificant increase with sleep loss. In contrast, the time taken for sending shows a progressive, significant increase with sleep loss. Also the number of errors made, but corrected, increases significantly with sleep loss. At 70 hr. of sleep loss, it is more than double the baseline period.

A concept-attainment task. This task, which is described fully in Bruner, Goodnow, and Austin (1956), requires the *S* to determine the properties that define class membership. The *S* faces an array of 81 cards on a board and these cards differ from one another in four respects: the figures on them vary in their shape (square, circle, or cross); their color (red, green, or black), their number (1, 2, or 3), and the lines bordering the card can be 1, 2, or 3 in number. The *S* is told that these cards can be grouped in different ways; e.g., all cards with two green crosses belong in a group, all other cards do not belong. The *E* then explains that he has a grouping in mind, and points out one card as belonging to this group. The *S* proceeds by choosing other cards one at a time. After each card, *S* is told whether or not the card belongs in the group and he continues until he can state what properties a card has to have in order to be in the group.

The *S*s were given six problems on each day of testing, with 5 min. per problem (a generous time limit). If the problem was not solved at the end of this time, *S* was told the answer, and then went on to the next problem.

Under these conditions, we find again no significant change in the general quality of performance. There is a slight decrease in the number of problems solved correctly (K is .12) and a small increase in the number of choices per solved problem (K is .22). In contrast, the time per solved problem shows a significant increase (K is .42). A general picture of results is presented in Fig. 5. All but two *S*s (out of twelve)[6] showed an increase in time, and after 77–81 hr. of sleep loss, these increases range from 16% to 71%. In contrast, all but one of the control *S*s required, on the same testing day, less time than they required at the end of the baseline session.

In short, on this task *S*s can again maintain their previous level of correctness, or a level very close to it, but they need more time.

Summary of complex S-paced tasks. On three tasks where it was possible to measure both speed and accuracy, changes in time but not in accuracy were consistently found. These tasks—addition, sending instructions, and concept attainment—differ in their

---

[5] Morning testing shows a similar slowing of work during sleep loss; K is .42 (.01 level), and upon the first recovery testing, the mean number attempted rises from 130 to 150. Percentage correct also shows some decline with sleep loss; K is .32 (.05 level), but the size of the change is small (from a baseline of 98.9% to 97.9% after 72 hr.). All results are drawn from the work of Nathene Loveland.

[6] In all, there were 16 experimental and 16 control *S*s, but time scores were not recorded for the first 8 *S*s.

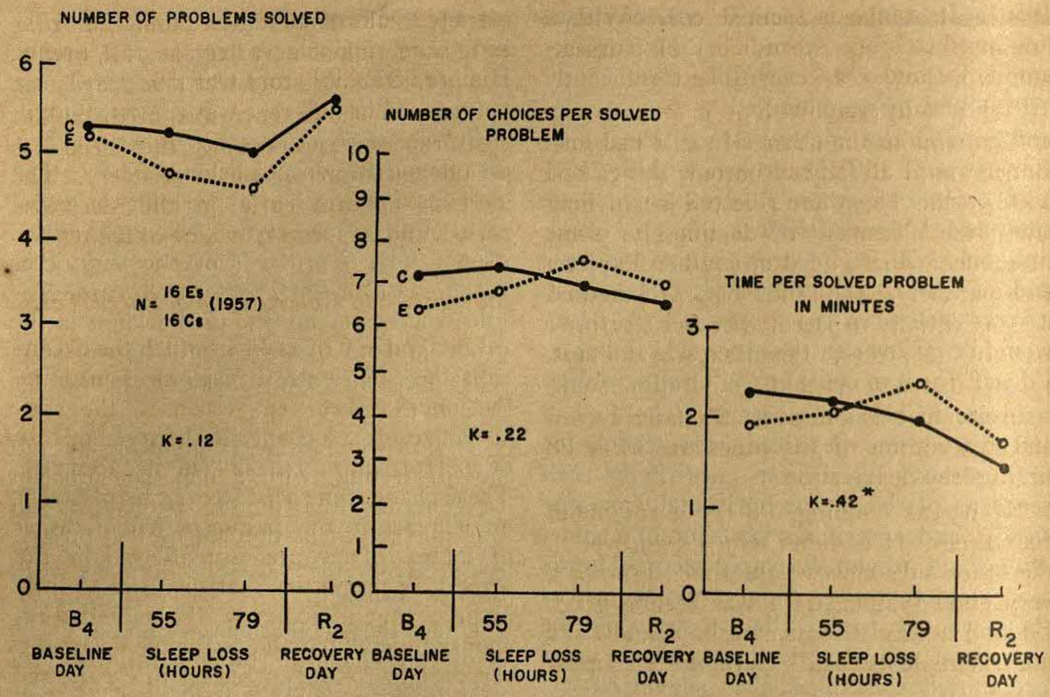


FIG. 5. Performance on a concept-attainment task. [United States Army photograph]

length and complexity but share two common properties. They are *S*-paced and the stress is on accuracy.

Changes in the Standard of Performance

In the previous three tasks, changes in speed can be expected simply on the grounds that lapses or brief periods of no response will occur. The lack of change in the quality of performance depends upon *S* having both the opportunity and the incentive to maintain accuracy by correcting errors or by keeping past events firmly in mind.

Suppose that this incentive was reduced or that the goal of the task could be changed by the *S*. In effect, the task would provide more than one level of success, and the *S* would control the level regarded as acceptable. A task in which the *S* predicts probable events provides for such a departure from the usual arrangement of either complete success or complete failure in solving a problem, and performance on it shows a decline in quality.

The *S* attempts to predict as accurately as possible whether a left- or a right-hand key will deliver a chip from a modified "slot-machine" (Goodnow, 1955). Success is set in terms of winning "around 50% or better" of the time. The *S* is playing against a random sequence of 144 events with the machine set to pay off 50% of the time on left and 50% of the time on the right. The *S* perceives the series as forming a pattern that is either very complex or subject to frequent change.[7] The variety of hypotheses about the pattern and the variety of attempts to "beat it" are limited only by the goal the *S* sets himself and his readiness or ability to keep in mind the past events that will reveal the pattern.

The major performance measure is the predictability of a choice from the outcome of the previous trial. This predictability is a measure of the variety of run-hypotheses that an *S* attempts (Hyman & Jenkin,

---

[7] This perception of a pattern is almost universal on a two-choice task, even when not fostered by *E* (Goodnow, 1955).

1956). Its range is from 0 to 1.[8] With a low predictability score like .04, for example, we find a statement like this:

"I'd usually stay with it if it had won and go over to the other side if it had lost. But it would all depend on how things had been going. Say there's been a lot of long runs and I figured it was time for some little ones. Then, I'd drop my usual system and switch around a bit more. If I figured it was going to go back and forth, I wouldn't go over to the other side if I lost. I'd wait for it to come up."

In contrast is the same *S*'s later record and description of his procedure after 28 hr. of sleep deprivation:

"If it was winning, I'd try it for about four times. If it didn't win, I'd go over to the other side most of the time. It's not a very good system. If I was feeling good, I'd find a way to outsmart it. I know it's not as simple as that." Here, the *S* adopted a simple rule of thumb without the "if's" and "buts" about a variety of past events. His predictability score was .56.

Figure 6 shows how this predictability measure increases during the sleep-loss period and drops again on recovery. The increase appears early in the sleep-loss period, and it is shown by nine of the ten *S*s.

Summary and Comment on S*-Paced Tasks*

On a variety of tasks in which the *S* controls the time when a response is made or the interval between responses, the consistent result is a change in the speed or rate of performance, but not in its accuracy. These changes in speed stem primarily from an increase in the frequency and duration of lapses. They are also affected by the same task properties that give rise to little or no change in accuracy: the unlimited response time which allows for the delay of a response or for the correction of errors, and the usual orientation toward correctness which prompts the *S* to sacrifice speed for accuracy.

The selection of any particular measure of time depends upon the task structure. In the simple reaction-time tasks that we have

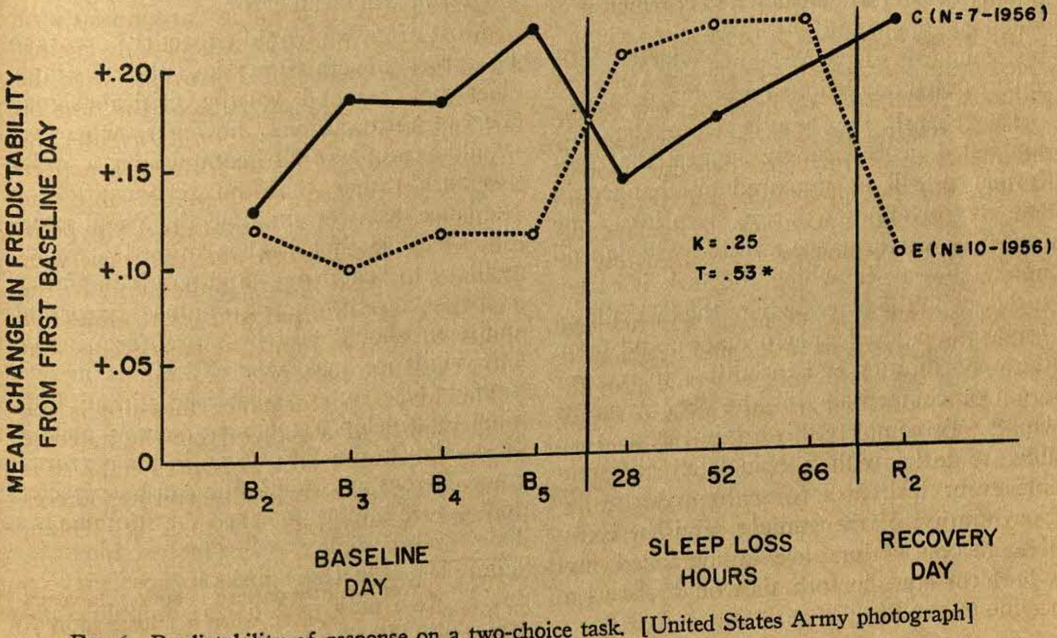


FIG. 6. Predictability of response on a two-choice task. [United States Army photograph]

used, where *E* controls the time when the stimulus appears, a regular reaction-time measure is appropriate, with the emphasis on changes in the range and distribution of these. In a serial reaction-time task, where the *S* determines the time when the next stimulus appears, delays between responses appear to be the most appropriate (Bjerner, 1949; Wilkinson, 1958). On more complex tasks like problem solving, the measures may be over-all time, output, corrected errors, bursts of work, or distribution of work over time, depending upon the way in which performance on the task is affected by lapses and the opportunities the *S* has to compensate for lapses.

---

[8] The score is the deviation from .50 of the proportion of times a correct choice is repeated, plus the deviation from .50 of the proportion of times that an incorrect choice is repeated. Additional details on results, apparatus, and procedure are given in Goodnow, Rubenstein, and Shanks (in press).

All of these suggested measures relate to time, and all of them are proposed for tasks where the *S* controls the pacing of responses. Changes on *S*-controlled tasks are not, however, necessarily limited to time. As Welford's (1953) analysis of fatigue indicates, errors may appear on *S*-paced tasks as a secondary form of deficit, reflecting perhaps the *S*'s anxiety about performance or his growing reluctance to make the effort demanded. Furthermore, changes in the quality of performance appear when the *S* controls the standards of performance. We observed, for example, a decline in the quality of performance on a task which provided an opportunity for the *S* to change his standards of performance where such a shift carried no obvious reminder of failure.

In general, the results from these *S*-controlled tasks strongly suggest that performance will be impaired or changed in one way or another during sleep loss. The aspect of performance affected will depend upon the way in which the task is structured, upon the places where the task allows for a relaxation of performance, and upon the opportunities and incentives to recover from or compensate for the occurrence of lapses. To complete this argument, we turn now to tasks with a completely different structure: tasks where such properties as the duration of the stimulus and the effective response time are *E*-controlled and where the lapse hypothesis would lead us to expect impairment in the form of errors.

EXPERIMENTER-PACED TASKS

In most *E*-paced tasks, the time available for an effective response is determined by *E* and is severely limited. The *E* may also control when the stimulus appears, the duration of the stimulus, or the required rate of performance. As in *S*-paced tasks, the response may be simple: in "vigilance" or "watch-keeping" tasks, for example, the *S* is required only to detect the occurrence of a critical signal. In other tasks, the response may be more complex: the *S* may be required to choose among several responses on the basis of information received or to commit to memory information available only briefly.

In *E*-paced tasks, either simple or complex, the lapse hypothesis leads us to expect an impairment in accuracy with sleep loss. The reasoning is as follows. Sleep loss will produce brief periods of no response in the face of stimuli which normally elicit a response, and these lapses will increase in frequency and duration as sleep loss progresses. In an *E*-paced task, the effective response time is limited. The signal in a vigilance task, for example, occurs briefly and without warning. If a lapse occurs at the time when a response should be made, the result will be an error.

Most often we would expect the errors to be errors of omission; i.e., the *S* simply does not respond to the critical signal. There are occasions, however, when we would expect errors of commission, a faulty response rather than no response. In a vigilance task, for example, the *S* will sometimes realize that he has missed a signal, will try to "catch up" and so give a faulty response on the next stimulus. Errors of omission should still be more frequent on any vigilance task where there is an extremely brief interval between stimuli, little time to develop a feeling of error, and no overt sign to the *S* of a gap in his performance. Errors of commission can be expected to be more important when the stimulus is a long and complex one (instructions, for example). On these tasks a lapse may coincide with only part of the stimulus, and, especially if some response is demanded, the effect will be a faulty rather than an omitted response.

Vigilance Tasks

In vigilance tasks, the *S* is required to make a simple response (usually pressing or releasing a lever) when and only when a critical signal occurs. In the present study, the signals varied in modality from task to task, and were either visual, auditory, or tactual. The aim of this was to ensure that errors were not simply the result of peripheral sensory difficulty (visual stimuli are especially suspect in this regard), and to check on the possibility that lapses were central in origin.

The visual task used the Continuous Performance Test (CPT) developed by Rosvold, Mirsky, Sarason, Bransome, & Beck (1956.)[9] For the first 10 min. the *S* responds by pressing a lever whenever the letter "X" appears in the viewer. (During the second 10 min. *S* responds to "X" only when it is preceded by "A." The results of the "AX" task will be analyzed in a later section.) In all, the X test covers 620 trials with the letter X occurring on 160 trials. The letters occur in a fixed sequence of 31 letters, in which X appears eight times. This sequence is repeated 20 times at the rate of approximately one letter per second; i.e., a total of approximately 10 min.

The auditory task was designed to duplicate the essential features of the visual task. A taped sequence of 30 letters was played to the S, who responded by pressing the lever when "X" was heard. As in the visual task, the rate was one letter per second and the total time was 10 min. The sequence was repeated 20 times, and the letter X occurred eight times in each sequence.

In the vibratory task,[10] the *S* released a signal key whenever he felt a vibration at either his left wrist or his right temple. All told, stimuli could come from any of six body areas, with vibrators (magnetic earphones) at each temple, wrist or knee. In one form of the task, only one area was stimulated at a time; in another, two areas were stimulated simultaneously. The interval between stimuli was again one second. Each sequence of 52 trials contained 18 critical stimuli and the sequence was repeated 12 times. In all tasks, responses were automatically recorded.

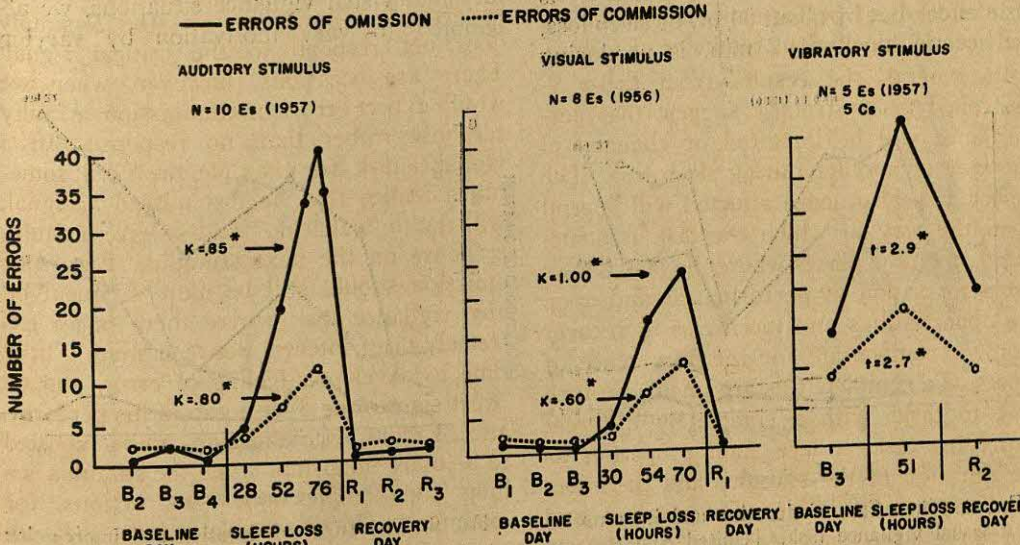


FIG. 7. The incidence of errors on three vigilance tasks. [United States Army photograph]

---

[9] We are grateful to Allen Mirsky for the loan of the apparatus.

[10] The vibratory task was designed and conducted by George H. Crampton. All other tasks in this section, with the exception of Schein's (1957) communication task, were administered by Charles F. Gieseking.

For illustrative purposes, the detailed results will be given for only one task in each modality. In all three of these tasks the usual condition of no knowledge of results prevailed.

Results. A picture of the size and nature of the decrement is given in Fig. 7 which shows errors of omission and commission for each sense modality. It is of interest that there are consistently more errors of omission than of commission and that omissions are more sensitive to sleep loss. The evidence that lapses occur in all three sense modalities strongly supports the notion that they are central in origin.

Conditions Affecting Decrement

The conditions with which we are concerned are much the same as those discussed in the section on *S*-paced tasks. Task duration is again important and the effect of task interruptions and changes in the stimulus conditions can be shown more clearly. Similarly, the conditions that are roughly called "motivation"—information about results, exhortations to do better, the challenge of a task—appear again to be less effective than task duration.

Task Duration. Figure 8 shows the incidence of errors in 2-min. blocks of the X and AX forms of the Rosvold-Mirsky CPT. On the first day of sleep loss there is little effect, but by about 54 hr. of sleep loss, the second block of 2 min. is producing more errors of omission than the first 2 min. At 70 hr. of sleep loss there is an even steeper rise in errors during the first 6 min., which then levels off.

Figure 8 also shows that the same effect occurs on Form AX, which followed Form X. It is interesting to note that during the first 2 min. of the AX task errors are much less frequent than during the last 2 min. of the X task. This brief recuperative effect may be due to either the short rest pause (1–2 min.) or to a slight change in task (from an "X" to an "AX" task), and the differential effects of these we cannot disentangle with the data we have. Combined, however, they come to play a more and more important part as sleep loss progresses.

As a check on these results with the Visual Vigilance Test, Fig. 9 shows that task duration has roughly the same effect on Forms X and AX of the Auditory Vigilance Test. In both tasks there is the strong implication that the important factor is not simply total time but the duration of an uninterrupted set of stimulus-response conditions. In both tasks there is also a definite interaction of task duration with sleep loss; as hours of sleep loss increase, the detrimental effect of task duration becomes stronger.

Effect of motivation. For both the vibratory and visual vigilance situations, we attempted to vary motivation by varying

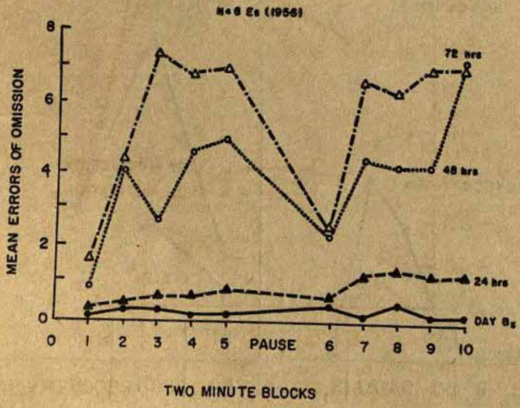


FIG. 8. The effect of duration and interruption on a visual vigilance task. [United States Army photograph]

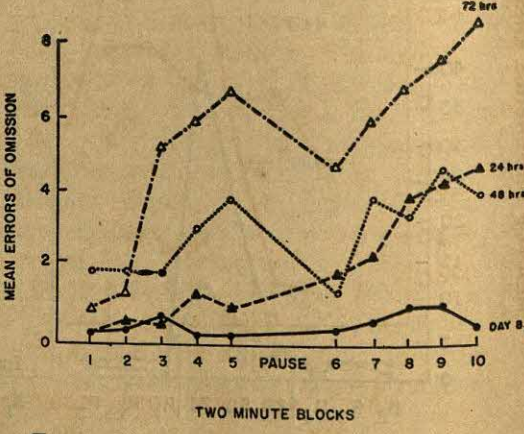


FIG. 9. The effect of duration and interruption on an auditory vigilance task. [United States Army photograph]

knowledge of results. In both tasks it was found that sleep-deprived *S*s with knowledge of results tended to show less impairment than sleep-deprived *S*s without knowledge of results, but the differences were small and somewhat inconsistent.

Effect of motivation on the Vibratory Vigilance Task. For one group of 12 *S*s, only one body area was stimulated at a time. In a second group of 12 *S*s, two body areas were stimulated simultaneously, a condition which makes discrimination of the critical stimuli more difficult. In each group, the six experimental and the six control *S*s were subdivided into high- and low-motivation conditions. Control and experimental *S*s were matched on baseline scores.

The high-motivation groups were exhorted to do better every day. A general knowledge of results was given; i.e., *S*s were not given absolute scores but were told whether they had performed better or worse on the previous test session and the percentage change in terms of their previous score. In contrast, the low-motivation groups were given no knowledge of results and no exhortation.

Difference scores (experimental minus control) were computed for each pair of *S*s on each test session, and the daily averages for the three pairs in each task are shown in Fig. 10. Although the high-motivation group shows less decrement during sleep loss, the differences between the high- and low-motivation groups are significant only for the single stimulus task. The average rank-order correlation, K, is .87 ($p < .01$) for the single stimulus experiment and .27 (n. s.) for the double stimulus experiment.

Effect of motivation on the Visual Vigilance Task, Form 2.[11] On a visual vigilance task, 20 *S*s were again divided into high-

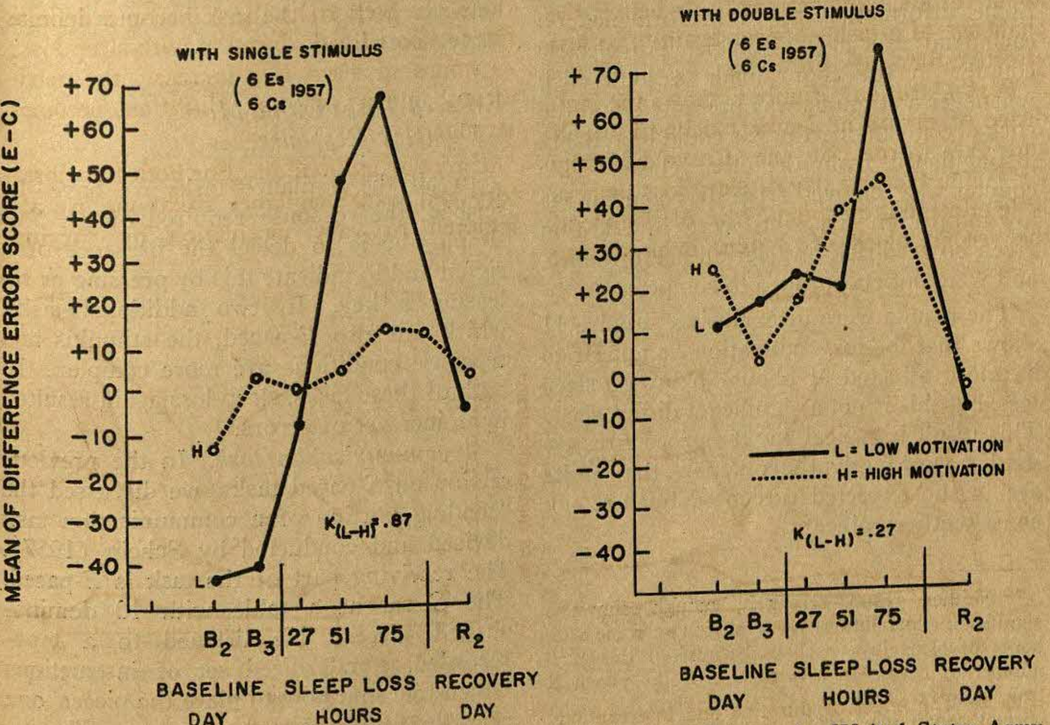


FIG. 10. Errors under high and low motivation on a vibratory vigilance task. [United States Army photograph]

---

[11] The apparatus for this task was modeled after the Rosvold-Mirsky CPT apparatus. The only differences are: (*a*) *S* released rather than pressed a lever when an X appeared, and (*b*) each letter was illuminated for only the half second during which it passed the aperture.

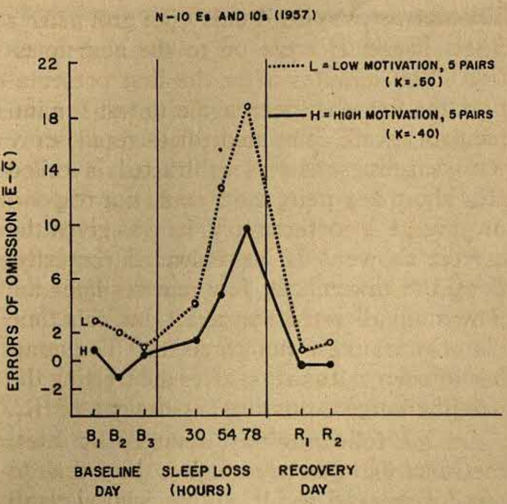


FIG. 11. Errors under high and low motivation on a visual vigilance task. [United States Army photograph]

and low-motivation groups. In the high-motivation condition, *S*s were given their scores at the end of each session, and were encouraged to compete with the other *S*s as well as with their own previous scores. Some information about performance on each trial was also available. An electric counter in the room made a loud click whenever a stimulus was presented, and the click accompanying the X's could be reliably distinguished from the other clicks. This counter was removed from the room for the low-motivation *S*s, who were in addition given no feedback on over-all performance, and were not encouraged to compete.

The results were inconclusive. Figure 11 shows that the low-motivation *S*s appear to be more affected at about 76 hr. of sleep loss, but this is not a significant difference.[12] This result is typical for the procedures we used to vary motivation: namely, the results are in the expected direction, but are not consistently significant.

Summary of vigilance tasks. Sleep loss consistently produced impaired performance on three vigilance tasks. On all three tasks, the major effect was an increase in errors of omission: the *S* missed increasingly often the occurrence of a critical signal. This is the result expected from lapses on a task where the critical stimuli occur briefly and without warning.

Over the three tasks, the modality of the stimulus was varied (stimuli were visual, auditory, or vibratory). The similarity of results on all three tasks makes it very unlikely that lapses are the result of any peripheral difficulty, and suggests instead that they are central in origin.

Of the two task conditions that were considered—task duration and knowledge of results—the former is again the more effective. Errors increase as a task continues without interruption and without a change in the stimulus-response conditions. As sleep loss progresses, errors begin to appear earlier and earlier in the task, and the benefit derived from an initial effort or a break between parts of the task becomes increasingly short-lived.

E-*Paced Tasks with a More Complex Response*

In all the vigilance tasks we have described, the response required of the *S* is simple—he is to detect the presence of a signal and to indicate this by pressing or releasing a key. In two additional tasks, which are also *E*-paced, the stimulus and response conditions are more complex. In both of these tasks, sleep loss again resulted in an increase of errors.

A communication task. In the previous section on *S*-paced tasks, we discussed the "sending portion of a communication task devised and conducted by Schein (1957). The receiving part of the task is *E*-paced. The *S* sat at a table with 10 domino-shaped pieces. He listened to a tape-recorded, standardized set of instructions which told him how to place the pieces, one at a time, to construct a pattern. The instructions were paced so that about 12 sec. were required to describe the location of a single piece and the instructions for each piece were prefaced by the number of the piece. During a given session *S*s completed five problems requiring 10 pieces each.

---

[12] Neither group, strangely enough, shows a significant correlation of omission errors with hours of sleep loss, perhaps because of the small number of cases. In a third test of the visual X task, however, with eight *E*s and eight *C*s under normal conditions, errors of omission were again significantly related to hours of sleep loss (K = .95*).

In this task the number of errors (incorrect placement of pieces) increased significantly with sleep loss. This is in contrast to the results for the sending (*S*-paced) part of the task where the significant decrement occurred in speed but not in accuracy. The errors are errors of commission, which makes sense in view of the possibility of a lapse coinciding with only part of the 12-sec. set of instructions for each piece, and in view of the fact that instructions give the number of each piece so that any gap in response can be detected by the *S* who must fill in the gap in some way before the next piece can be used.

An information-learning task. The stimuli were short items of information, and the response required of the *S* was to commit these to memory. The task was *E*-paced in that the information was available only for a brief period of time. We would expect, with sleep loss, an increasing number of items to be "missed," i.e., not learned to the point of being available for recall. This is, in fact, the major result obtained.

A list of some 428 information items was culled from various sources such as *Information Please Almanac*. From this pool, 275 items were selected. We tried to pick out items that would be of general interest, but not generally known. For example: "Who was the first President to die in office?" By a random procedure, 25 items at a time were selected from the 275 to make up 11 lists, one for each day of the experiment.

On the screening day, each *S* was pretested on the 11 lists so that a baseline score could be obtained. Answers were not given during this pretest. The control *S*s were matched individually with the experimental *S*s on this baseline score.

During each session, while the *S* stood with his eyes closed having his EEG recorded, *E* read aloud the 25 items and their answers. He read at a prescribed rate, pausing 5 sec. between question and answer. The answer was read only once and after a 5-sec. pause, *E* went on to the next question. Ten minutes after the first presentation, the list was given again to test for immediate recall. The immediate recall score with baseline answers subtracted is called R_1. If, on any item, the *S* could not respond or gave an incorrect reply, he was given the correct answer. If he answered correctly, *E* said "Correct." A few minutes later the *S* was again tested for recall, but this time his answers were not corrected. The number of correct answers, after subtracting the baseline correct answers, is designated R_2.

On the following day, some 23 hr. later, the same list was presented a third time to test for delayed recall. (This delayed recall score, with baseline score subtracted, is called R_3.) The *S*'s answers were corrected. Later in that session, a new list was presented, and so on, with a different list each day. On any one day all *S*s received the same list.

Results. Figure 12 shows the results with 10 experimentals and 10 matched controls, for the initial recall score, R_1. This score, R_1, is essentially the equivalent of correct responses on a vigilance task, the score declining as more and more questions are missed (responded to mostly with the answer "Don't know" or "Don't remember"). The sleep-deprived *S*s show impairment at about 28 hr. of sleep loss, and the decline continues throughout the 76 hr. of sleep loss (K = .67*). There is a signifi-

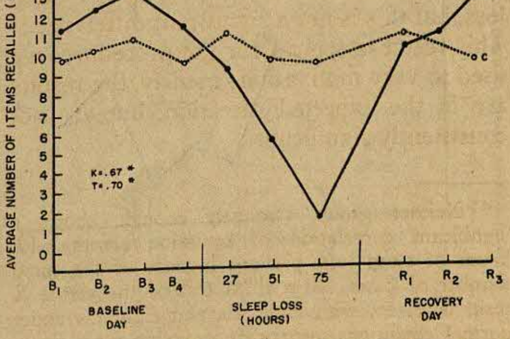


FIG. 12. The effect of sleep loss on immediate recall. [United States Army photograph]

cant recovery immediately after the first sleep. The improvement continues and the scores return to about the control level after 25 hr.[13]

A failure in immediate acquisition or immediate recall is easily accounted for by a lapse hypothesis. The *S* simply fails to take in information that is only briefly available. This failure, however, is not the only important facet of a learning task. For these other facets, the scores R_2 and R_3 are relevant. R_2 (recall a few minutes after R_1) declines but is more resistant to the effects of sleep loss than R_1, which is not surprising since R_2 follows a few minutes after active questioning and providing of correct answers on the occasion R_1. Of most interest is the effect of sleep loss on delayed recall.

Delayed recall during sleep loss (score R_3 minus score R_2) is shown in Fig. 13. Recall after 24 hr. (R_3) is as accurate as recall after a few minutes (R_2) for the first day of sleep loss. After 51 hr. of sleep loss have gone by, however, delayed recall is markedly poorer and remains so until the recovery period. Thus, this loss does not increase steadily as sleep loss increases (K = 0), but scores are, in general, lower during sleep loss than during baseline and recovery periods (T = .70*). This change in sleep loss cannot be accounted for by a lapse hypothesis. If the results could be substantiated, an additional concept would be necessary. Unfortunately, we have only the one task in which a change in delayed recall could appear.

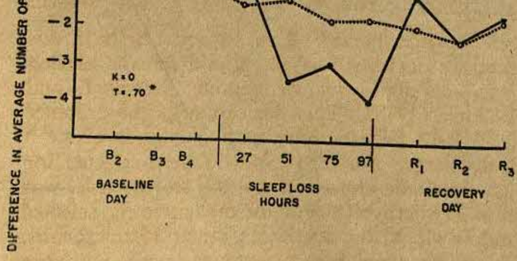


FIG. 13. The effect of sleep loss on delayed recall. [United States Army photograph]

Summary and Comment on E*-Paced Tasks*

It may seem strange that we put together a communication task, a learning task, and a number of vigilance tasks, without regard for the mental processes assumed to be involved. In all of these tasks, however, the effective response time is brief and controlled by *E*, and in all of them the results substantiate the expectation that in *E*-paced tasks impairment with sleep loss will take the form of an increase in errors. For *S*-paced tasks, it will be recalled, changes in speed rather than in accuracy were expected.

The incidence of errors was found to be mostly affected by the length of time during which a task continued without interruption and without a change in the stimulus response conditions. Task duration in this sense interacts with sleep loss, so that with increasing hours of sleep loss, errors start to occur earlier and earlier in the task. The other condition considered—knowledge of results—showed only a slight and somewhat inconsistent effect on performance. These results are the same as those found with *S*-paced tasks.

At this point, we feel that the usefulness of the lapse hypothesis has been amply demonstrated. It offers a very fruitful pointer to the aspects of performance in which impairment will appear, and is a guide to the scores or measures that will be most sensitive to the effects of sleep loss. In addition, the occurrence of lapses in three sensory modalities and a large variety of tasks suggests that lapses are a central rather than a peripheral phenomenon.

With this last point in mind, we turn to material on changes in the EEG alpha rhythm with sleep loss and to relationships between these changes and the occurrence of lapses. This material is particularly relevant to the identification of lapses as transient states of sleep or a condition close to sleep.

---

[13] In all of the statistical analyses we used the difference between the scores of the experimental *S* and his matched control. This procedure compensates for practice effects and differential difficulty of lists.

AN ELECTRO-PHYSIOLOGICAL CORRELATE OF IMPAIRMENT[14]

In the history of sleep loss, there have been a number of attempts to find impairment in physiological functions. Of these, the most consistent positive result is a decline in the EEG alpha rhythm with eyes closed (Bjerner, 1949; Kleitman, 1939; Malmo, 1958; Tyler et al., 1947). Changes in alpha rhythm also have been shown to be related to brief changes in performance.

In Bjerner's 1949 study of sleep loss, for example, depression of alpha activity was significantly associated with "delayed actions" on a reaction-time task. In a later study, Simon and Emmons (1956) found a supporting result with sleepy or sleeping *S*s: as the quality and quantity of alpha increased, so did the probability that a stimulus would be reported heard and correctly recalled later. The stimuli were answers to information questions.

These findings strongly suggest that changes in alpha rhythm are a likely place to find a physiological correlate of impairment and lapses. In this section we present part of the material gathered by Armington and Mitnick (1959) substantiating a general decline in alpha with sleepiness, together with the results of an auditory vigilance task with simultaneous EEG recording. On this task, errors of omission were significantly associated with lower alpha activity. It will be recalled that, in an *E*-paced task like vigilance, errors of omission reflect the occurrence of lapses. We are strongly inclined to follow Bjerner's argument that both the general decline in alpha with sleep loss and the more transient declines accompanying brief impairments of performance signal a state that is similar to sleep itself.

Sleep loss and a decline in alpha amplitude. As a start, let us first establish that in the present study, alpha with eyes closed declines with sleep loss. The decline is clearly shown in Fig. 14, taken from work by Armington and Mitnick (1959).[15]

In Fig. 14, alpha during sleep loss is considerably less when the *S* is instructed to "keep his mind blank" or to count from one to ten than when he is instructed to add 17's. This is a reversal of the relationship found in the normal awake state (control period). A similar reversal has been noted before with fatigue (see Bjerner, 1949, p. 74) and with narcolepsy (Daly & Yoss, 1957). The result gives strong support to the frequently advanced hypothesis that a decline in alpha occurs with a shift, in either direction, away from an intermediate level of alertness, tonus, excitation, or mental activity (cf. Bjerner, 1949; Lindsley, 1951; Stennett, 1957). In a normal awake state, a shift upwards and a depression of alpha can be induced by mental work or by a startle stimulus. During sleep loss, the depression of alpha is induced by a shift down from the intermediate level; in this state the introduction of mental work or a startle stimulus shifts the *S* away from a drowsy state, back toward the intermediate level of activity and the reappearance of alpha.

Alpha and errors of omission. We wish to compare the presence of alpha during errors of omission with the amount present when other responses are made. With this aim in mind, the response to each stimulus on the auditory vigilance task was classified as follows: (*a*) correct response—the critical signal (the letter X) occurred, and *S*

---

[14] For the material in this section, the guidance and help of John Armington was invaluable.

[15] For Fig. 14 and all subsequent results in this section, scores were based on an automatic scoring device developed by Armington, Biersdorf, and Mailloux (1958). This device marks the presence of all 8–12 CPS activity with amplitude above a fixed voltage. For eight *S*s who were selected for high-amplitude alpha activity, the scoring level was set at 25 microvolts. For the other two *S*s, selected for low alpha, the level was set at 15 microvolts. For all *S*s, bipolar recording was used, with one electrode on the vertex and the other on the occipital prominence.

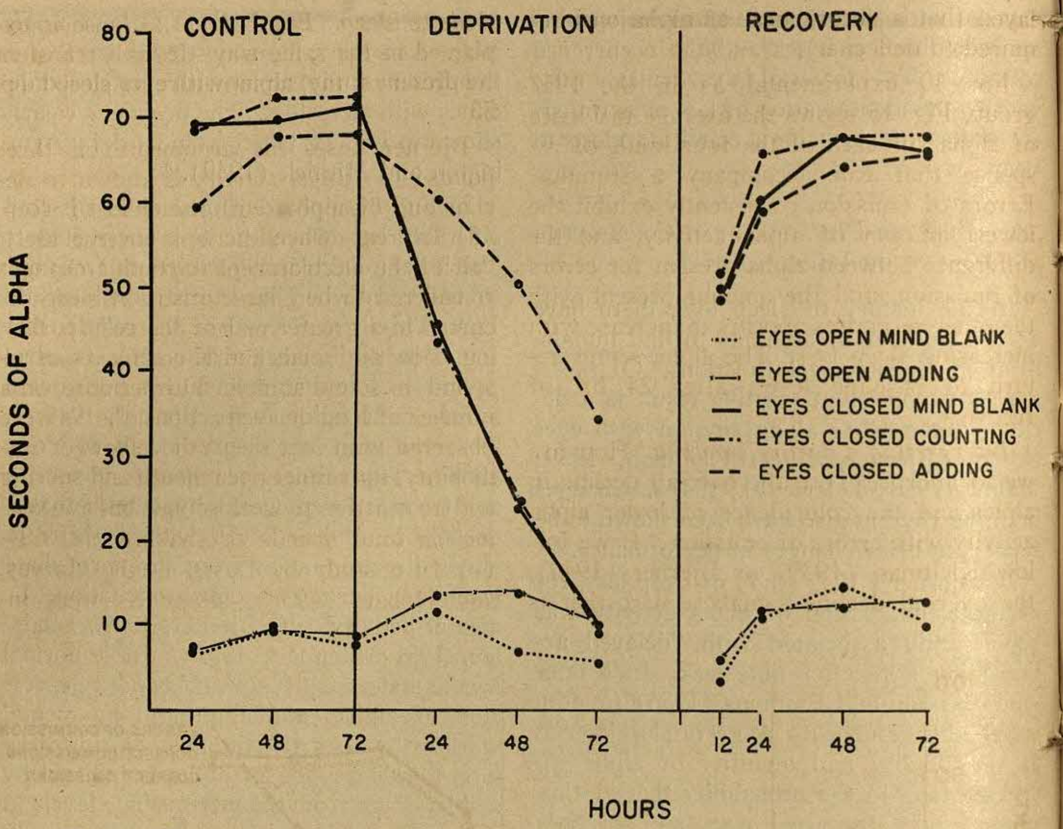


FIG. 14. Changes in alpha under several task conditions. [United States Army photograph]

responded with an appropriate release of the lever; (*b*) correct withholding of response—a letter other than X occurred, and *S* did not release the lever; (*c*) error of commission—the *S* released the lever for a letter other than X; (*d*) error of omission —the letter X occurred, and *S* failed to release the lever.

In determining the presence of alpha for each of these four types of response, the immediate problem is one of determining the time period around each response during which alpha could be counted. Simon and Emmons (1956) used the 2–7 sec. interval (average was 4 sec.) during which the answer to the information question was given to *S*. Bjerner (1949) used the time during which the delayed action occurred (most of these were 2–3 sec.), together with the 3 sec. before and after each delayed action. We have followed Bjerner in using a before-during-and-after interval, but have narrowed the interval down to 1 sec. before and 1 sec. after each stimulus letter. In our vigilance task, stimuli occur at the rate of one letter per second, and a longer interval gives rise to the problem that a single interval sometimes contained two errors of omission.[16]

For each of these brief intervals around each stimulus and response, alpha was scored as present or absent. Alpha was scored as present if there was any deflection of the automatic scoring pen (to avoid response to noise, the pen's action was so delayed that a short burst of alpha was required to deflect it).

---

[16] For one *S*, as a trial procedure, we used three time intervals to compare differences in alpha activity for the several responses. These intervals were: (*a*) from 1 sec. before to 1 sec. after the stimulus (results significant at .001 level), (*b*) over the 4 sec. before a stimulus (significant at .05 level), and (*c*) over the 2 sec. before a stimulus (significant at .01 level).

For 10 experimental *S*s in the 1957 group, Fig. 15 shows the average incidence of alpha for each of the four kinds of response that can accompany a stimulus. Errors of omission consistently exhibit the lowest amount of alpha activity, and the difference between alpha present for errors of omission, and the amount present with the other responses appears to increase with increasing sleep loss. The alpha scores return to baseline levels after 24 hr. of recovery.

Interpreting a decline in alpha. How are we to interpret both the over-all decline in alpha and the coincidence of lower alpha activity with errors of omission? If we follow Kleitman (1939) or Bjerner (1949), the over-all decline signals a state that is close to sleep. For Bjerner, a lapse is explained in the same way. It "is a transient condition of the same nature as sleep" (p. 73).

Bjerner bases his arguments on three points: (*a*) alpha activity is known to decline and disappear with the onset of sleep. (*b*) In long delayed actions (over 5 sec.), "all of the electroencephalographic changes considered to be characteristic of sleep occurred to a greater or less degree" (p. 68), e.g., slow potentials and K-complexes in response to sound stimuli. Furthermore, on a number of long delayed actions, the *S*s were observed to in fact sleep: they showed "immobility, sometimes open mouth and snoring and no reaction to weak sounds but awakening on loud sounds or shaking" (p. 68). (*c*) In a study by Davis, Davis, Harvey, and Hobart (1938), sleepy *S*s were in-

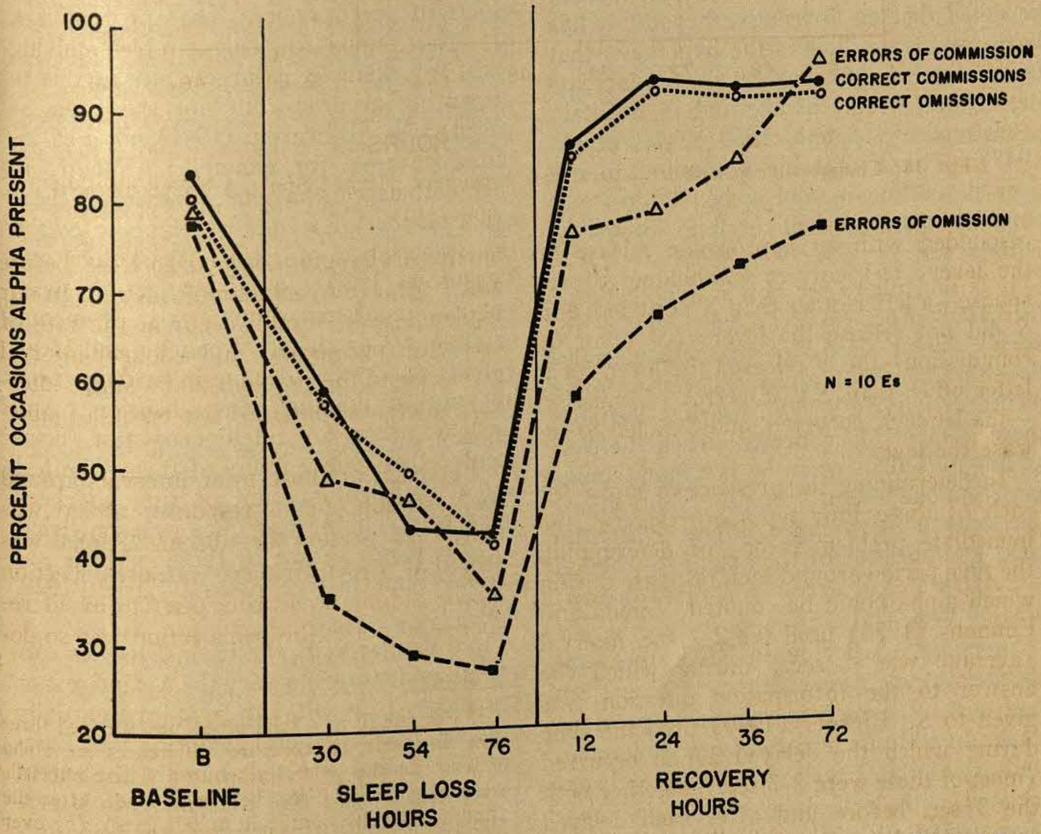


FIG. 15. The incidence of alpha with errors and correct responses on a vigilance task. [United States Army photograph]

structed to give a signal whenever they felt a spell of drowsiness. Most of these signals were preceded by depressions of alpha.

Additional support for regarding the decline in alpha amplitude as a sign of a condition like sleep was found in the present study by Armington and Mitnick (1959). The wave-form of the EEG was observed to be similar to that described by Simon and Emmons (1956) and by Jung (1954) as typical of drowsiness or of the transition state between wakefulness and sleep. On this basis, the decline would be expected to be greatest on a task which facilitates drowsiness ("mind blank" or counting from one to ten), and less on a task which arrests drowsiness (for example, adding 17's). These are, in fact the results presented in Fig. 14.

Although these various results present a consistent picture of changes in alpha occurring with a condition like sleep, it must be noted that an alternative hypothesis has been proposed. This is the hypothesis that the decline of alpha reflects the *S*'s increasing efforts to stay awake and to maintain performance (Malmo, 1958; Tyler et al., 1947). The hypothesis is limited to the over-all decline in alpha; i.e., it does not consider the transient declines occurring with lapses.

It is difficult to see how a hypothesis of increased effort could account for the reversal, during sleep loss, of the usual relationship between alpha depression and mental activity (Fig. 14). During sleep loss there is more alpha present when the *S* is adding 17's than when he is simply counting from one to ten. It appears more plausible to us to interpret the difference between tasks as indicative of the extent to which a task facilitates or arrests a state of drowsiness or near-sleep.

Summary. A decline in EEG alpha amplitude was observed with increasing sleep loss, a result which substantiates previous work. In addition, errors of omission on an auditory vigilance task were consistently associated with less alpha than were other responses. Both of these findings are interpreted as indicating that in sleep loss generally, and during lapses particularly, the *S* is in a state similar to or close to sleep itself.

Discussion

With the results and the ideas presented, we had hoped: (*a*) to account for the apparently inconsistent effects of sleep loss, (*b*) to provide some guide toward effective testing for impairment, and (*c*) to make a start on describing the general nature of impairment, and relating psychological to physiological changes.

As a way of accounting for negative and inconsistent results, the relationships between lapses and the properties of a task are decidedly useful. They enable us to draw together within the present study the results of a large variety of tests. Moreover, when we look back at previous studies, negative results can often be explained in terms of failure to measure the sensitive aspect of performance. The apparent lack of effect upon difficult problem-solving tests, for example, is largely attributable to measuring accuracy but not speed on *S*-paced tasks (Edwards' [1941] untimed intelligence test, for example). Where, in fact, both aspects have been measured, there are a number of results which point toward the importance of time. Weiskotten and Ferguson (1930) found changes in speed but not in accuracy on a task of mental arithmetic. Carmichael, Kennedy, and Mead (1949) observed a decline in reading speed but not in reading comprehension. And Laslett, using an intelligence test, commented that "Since many of them [the *S*s] had a sufficient amount of time to check their work on the test once or even twice, . . . the losses shown by these subjects . . . are less than the true losses" (Laslett, 1928, p. 391).

In a similar vein, the inconsistencies surrounding attempts to vary the *S*'s interest or knowledge of results can be at least partially resolved by the general condition that Bartlett (1942) has stressed: namely, the extent to which the *S* relaxes his standards of performance. With this general condition, attention focuses on the way in which any particular introduction of information, reminders, warnings, or exhortations counteracts the relaxation of standards, and shifts away from a provocative but rather unwieldy condition like motivation or arousal (Ax et al., 1957; Wilkinson, 1957).

To take the second aim, how effective a guide in testing for impairment are the ideas and results presented? In the first place, they run counter to the arguments that, since test performance shows so little change, we should concentrate on changes in behavior and mood (Kleitman, 1939) or on the measurement of effort (Tufts College, 1949). In fact, it now seems doubtful to us that the much-discussed changes in mood and disposition (Kleitman, 1939; Tyler et al., 1947) are directly attributable to sleep loss, either by way of a loss of judgment or a lowered threshold of irritability. In the present study and in the study by Eagles et al. (1953), incidents of anger and irritability were rare. In both studies, when anger appeared it was in resentment of a particular method of being kept awake or a face-saving denial of having fallen asleep. It seems quite likely that disposition changes are primarily a function of the *S*'s attitudes and the methods of control rather than any direct effect of sleep loss.

More specifically, the relationships between lapses and the properties of a task provide some very definite guides to the aspects and measures of performance which are most likely to be sensitive. In this respect, there is some guarantee against the conclusion of "no impairment" which often follows a failure to measure the critical aspect. Furthermore, the conditions found to promote impairment—task monotony and the opportunity to relax standards—suggest ways of making the appearance of impairment more likely.

There is, however, a major gap in the concepts presented. This lies in the relationship between lapses and performance on tasks which are more automatic and less dependent on stimulus discrimination than the ones we have used. Singleton (1953), for example, using a task where perceptual and motor speed could be measured separately, found that only the perceptual part of performance was affected. And Wilkinson (1958), using a vigilance task where the lapse hypothesis would lead one to concentrate on errors of omission or nonresponse to a stimulus, found that errors of commission become more important if a rhythmic sequence of responses was set up.

The difficulty in handling these tasks seems to lie, as Broadbent (1953b) has suggested, in the definition of a lapse as a period of no response. It may be more appropriate to regard lapses as periods in which the *S* is less responsive to a stimulus, and in which the gap may be appropriately filled by a response which is highly practiced and less dependent on stimulus discrimination, or inappropriately filled by a competing response. This type of definition would fit with the analyses of motor tasks that have stressed the disorganization or dislocation of units in performance (Bartley & Chute, 1947; Welford, 1953), and would also fit with Bills' (1935, 1937) finding that blocks are promoted by the presence of competing responses and lessened by practice. Some such additional definition would clearly be needed before the lapse hypothesis could be used with profit for tasks where the response hinges less upon the discrimination of a stimulus. For the tasks used in the present study, however, the definition of a lapse as a period of no response is adequate.

The final question, and the most difficult, is the general nature of impairment. Up to a point, it is extremely helpful to view impairment during sleep loss in terms of lapses which reflect brief periods of sleep or a condition like sleep. This approach, for example, makes it possible to obtain very clear relationships between performance on psychological tests and changes in alpha rhythm.

The need to extend this view of impairment comes from two points: (*a*) there is as yet no clear relationship between the nature of impairment and the conditions promoting impairment, and (*b*) lapses have been found to characterize impairment under conditions which do not appear to be the same as sleep loss, namely hypoxia (Maag, 1957) and psychosis (Eysenck, Granger, & Brengelmann, 1957, pp. 100–101).

On the first of these two points, we can do little more than point to the problem and some suggested solutions. At an empirical level, for example, there is no difficulty in stating that impairment is promoted by an uninterrupted and unchanging set of stimuli and responses, both in studies of sleep loss and in studies of fatigue. The problem lies in describing this effect in less empirical terms.

It is this problem which is the core of Solomon's (1948) emphasis on stimulus satiation in prolonged work, and of Broadbent's (1957) model for perception and memory in terms of stimulus satiation and stimulus sensitivity. It is also the heart of Kleitman's (1939) view of wakefulness as maintained by incoming stimuli and accordingly decreased either by monotonous stimuli or by the lesser stimulation that goes with physical immobility. None of these approaches offers a very specific guide as to the optimal arrangement of stimuli and responses, but the proposals of Solomon (1948) and Broadbent (1957) do suggest some possible and needed lines of research.

What are we to make of the finding that lapses characterize impairment under conditions like hypoxia (Maag, 1957) and psychosis (Eysenck et al., 1957)? These results make it impossible to regard lapses as always reflecting brief periods of a condition like sleep. What appears far more likely, and indeed more encouraging, is that impairment under any condition takes the form of an increasing irregularity or unevenness of performance (cf. Maag, 1957). For the specific condition of sleep loss, the major contributor to the unevenness is the occurrence of brief periods of sleep. From this point of view, the contributing factors may vary from one impairing condition to another, but performance itself always contains the possibility of unevenness. This feature of performance has been well described by Broadbent: "Crudely speaking, a man is not like a child's mechanical toy which goes slower as it runs down. Nor is he like a car engine which continues normally until its fuel is exhausted and then stops dead. He is like a motor which after much use misfires, runs normally for awhile, then falters again, and so on" (Broadbent, 1955, p. 2). This description is as apt for performance under sleep loss as it is for performance with prolonged work.

SUMMARY

With 49 *S*s, deprived of sleep for 72–98 hr., performance deteriorated on a variety of tasks, an unusual result in studies of sleep loss. Deficit took the form of lapses (brief periods of no response accompanied by extreme drowsiness and a decline in EEG alpha amplitude). Four features of lapses were noted. (*a*) They occur in other conditions such as fatigue and hypoxia and appear to characterize impairment in general. (*b*) They increase in both frequency and duration as sleep loss progresses. (*c*) They are strongly affected by stimulus monotony. (*d*) Their specific effect on performance varies with the properties of the task. In *S*-paced tasks, for example, speed is the critical measure; in *E*-paced tasks, errors are critical. To identify the sensitive aspect of performance becomes the crucial problem.

REFERENCES

Armington, J. C., Biersdorf, W. R., & Mailloux, L. D. An electroencephalogram recording apparatus. *Amer. J. Psychol.,* 1958, **71,** 594–599.

Armington, J. D., & Mitnick, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.* 1959, **14,** 247–250.

Ax, A. F., Fordyce, W., Loovas, I., Meredith, W., Pirojnikoff, L., Shmavonion, B., & Wendahl, R. Quantitative effects of sleep deprivation. *USA Qm. Res. Develpm. Res. Rep.,* 1957.

Bartlett, F. C. Fatigue following highly skilled work. *Proc. Roy. Soc.,* 1942, **131,** 247–257.

Bartley, S. H. Fatigue and inadequacy. *Physiol. Rev.,* 1957, **37,** 301–324.

Bartley, S. H. & Chute, E. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.

Bills, A. G. Blocking: A new principle in mental fatigue. *Amer. J. Psychol.,* 1931, **43,** 230–245.

Bills, A. G. Some causal factors in mental blocking. *J. exp. Psychol.,* 1935, **18,** 172–185.

Bills, A. G. Fatigue in mental work. *Physiol. Rev.,* 1937, **17,** 436–453.

Bjerner, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test: A study in sleep deprivation. *Acta Physiol. Scandinavica,* 1949, **19,** Suppl. 65. Pp. 1–93.

Broadbent, D. E. Neglect of the surroundings in relation to fatigue decrements in output. In W. F. Floyd & A. T. Welford (Eds.), *Fatigue.* London: Lewis, 1953. Pp. 173–178. (a)

Broadbent, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.,* 1953, **44,** 295–303. (b)

Broadbent, D. E. Variations in performance arising from continuous work. In *Conference on individual efficiency in industry.* Cambridge, England: Med. Res. Council, 1955. Pp. 1–5.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.,* 1957, **64,** 205–215.

Bruner, J. S., Goodnow, Jacqueline J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Carmichael, L., Kennedy, J. L., & Mead, L. C. Some recent approaches to the experimental study of fatigue. *Proc. Nat. Acad. Sci.,* 1949, **35,** 691–696.

Daly, D. D., & Yoss, R. E. Electroencephalogram in narcolepsy. *EEG clin. Neurophysiol.,* 1957, **9,** 109–120.

Davis, H., Davis, P. A., Harvey, E. N., & Hobart, G. Human brain potentials during the onset of sleep. *J. Neurophysiol.,* 1938, **1,** 24–38.

Eagles, J. B., Halliday, A. M., & Redfearn, J. W. T. The effect of fatigue on tremor. In W. F. Floyd & A. T. Welford (Eds.), *Fatigue.* London: Lewis, 1953. Pp. 41–58.

Edwards, A. S. Effects of the loss of 100 hours of sleep. *Amer. J. Psychol.,* 1941, **54,** 80–91.

Eysenck, H. J., Granger, G. W., & Brengelmann, J. C. *Perceptual processes and mental illness.* New York: Basic Books, 1957.

Goodnow, Jacqueline J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.,* 1955, **68,** 106–116.

Goodnow, Jacqueline J., Rubenstein, I., & Shanks, Betty L. The role of past events in problem solving. *J. exp. Psychol.,* in press.

Hyman, R., & Jenkin, N. S. Involvement and set as determinants of behavioral stereotypy. *Psychol. Rep.,* 1956, **2,** 131–146. Monogr. Suppl. 3.

Jonckheere, A. R. A distribution-free K-sample test against ordered alternatives. *Biometrika,* 1954, **41,** 133–145.

Jung, R. Correlation of bioelectrical and autonomic phenomena with alterations of consciousness and arousal in man. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 310–339.

Kleitman, N. *Sleep and wakefulness.* Chicago: Univer. Chicago Press, 1939.

Laslett, H. R. An experiment on the effects of loss of sleep. *J. exp. Psychol.,* 1924, **7,** 45–58.

Laslett, H. R. Experiments on the effects of the loss of sleep. *J. exp. Psychol.,* 1928, **11,** 370–396.

Lee, M. A., & Kleitman, N. Studies on the physiology of sleep: II. Attempts to demonstrate functional changes in the nervous system during experimental insomnia. *Amer. J. Physiol.,* 1923, **67,** 141–151.

Lindsley, D. B. *Emotion.* In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

Lyerly, S. B. The average Spearman rank correlation coefficient. *Psychometrika,* 1952, **17,** 421–428.

Maag, C. H. Characteristics of mental impairment in hypoxia. *Amer. J. Psychol.,* 1957, **70,** 243–247.

Mackworth, N. H. Researches on the measurement of human performance. *MRC Spec. Rep. No. 268.* London: Her Majesty's Stationery Office, 1950.

Malmo, R. B. Certain physiological correlates of psychomotor functioning. Office of the Surgeon General, *USA Res. Rep.,* Contract No. DA-49-007-MD-626, 1958.

Murray, E. J., Schein, E. H., Erikson, K. T., Hill, W. F., & Cohen, M. The effects of sleep deprivation on social behavior. *J. soc. Psychol.,* 1959, **49,** 229–236.

Murray, E. J., Williams, H. L., & Lubin, A. Body temperature and psychological ratings during sleep deprivation. *J. exp. Psychol.,* 1958, **56,** 271–273.

PATRICK, G. T. W., & GILBERT, J. A. On the effects of loss of sleep. *Psychol. Rev.*, 1896, **3**, 469–483.

RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214–225.

ROSVOLD, H. E., MIRSKY, A. F., SARASON, I., BRANSOME, E. D., JR., & BECK, L. H. A continuous performance test of brain damage. *J. consult. Psychol.*, 1956, **20**, 343–350.

SCHEIN, E. H. The effects of sleep deprivation on performance in a simulated communication task. *J. appl. Psychol.*, 1957, **41**, 247–252.

SIMON, C. W., & EMMONS, W. H. EEG, consciousness, and sleep. *Science*, 1956, **124**, 1066–1069.

SINGLETON, W. T. Deterioration of performance on a short-term perceptual-motor task. In W. F. Floyd & A. T. Welford (Eds.), *Fatigue*. London: Lewis, 1953. Pp. 163–172.

SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.

STENNETT, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG clin. Neurophysiol.*, 1957, **9**, 131–138.

TUFTS COLLEGE, INSTITUTE FOR APPLIED EXPERIMENTAL PSYCHOLOGY. *Handbook of human engineering data.* Pt. 7. Medford, Mass.: Tufts Coll., 1949.

TYLER, D. B. Fatigue of prolonged wakefulness. *Nat. Res. Council Res. Rep.*, 1946, OE Memr-282, Part 2.

TYLER, D. B., GOODMAN, J., & ROTHMAN, T. The effect of experimental insomnia on the rate of potential changes in the brain. *Amer. J. Physiol.*, 1947, **149**, 185–193.

WARREN, N., & CLARK, B. Blocking in mental and motor tasks during a 65-hour vigil. *J. exp. Psychol.*, 1937, **21**, 97–105.

WEISKOTTEN, T. F., & FERGUSON, J. E. A further study of the effects of loss of sleep. *J. exp. Psychol.*, 1930, **13**, 247–266.

WELFORD, A. T. The psychologist's problem in measuring fatigue. In W. F. Floyd & A. T. Welford (Eds.), *Fatigue*. London: Lewis, 1953. Pp. 183–192.

WILKINSON, R. T. Effects of lack of sleep. Cambridge, England: Med. Res. Council, *FPRC Rep. No. 961.3*, 1957.

WILKINSON, R. T. The effects of sleep loss on performance. Cambridge, England: Med. Res. Council, *APRU Res. Rep. No. 323*, 1958.

(Received October 11, 1958)

Vol. 73, No. 15 Whole No. 485, 1959

Psychological Monographs: General and Applied

PERSONALITY AND ENVIRONMENTAL FACTORS IN THE DEVELOPMENT OF DELINQUENCY [1]

ROBERT D. WIRT AND PETER F. BRIGGS

University of Minnesota

THE VARIOUS PROBLEMS associated with the phenomenon called *delinquency* have been considered and discussed widely in the recent past. The conclusions to be drawn from the research literature in this area are variable. There are studies (e.g., Landers, 1954; Loevinger, Gleser, & DuBois, 1953; Shaw & McKay, 1942; Sullinger, 1936; Wier, 1944) definitely associating the incidence of delinquency with certain social conditions, subcultural areas, and environmental and family background. There are others (e.g., Briggs, 1955; Cleckley, 1941; Glueck & Glueck; 1934, 1940, 1950; Johnson & Szurek, 1949; Quay & Peterson, 1958) which show very clear relationships between certain personality traits and delinquent behavior. The present investigation began with a question which combines these two approaches: Given certain personality characteristics, which social factors are influential in directing an individual toward a path of delinquency?

THE PROBLEM OF THE DEFINITION OF DELINQUENCY

There are several ways of defining delinquency and its relation to antisocial acts. There is the ethical, moral, or religious point of view in which perhaps all behavior in violation of values could be considered delinquent. For the most part, this view of delinquency is not amenable to empirical study, important though it may be in our social structure.

---

[1] This investigation was supported by research grant Nos. M-1151 and M-1151(C) from the National Institute of Mental Health; United States Department of Health, Education, and Welfare. We wish to express our deep appreciation to the government for its generous financial support and to the University of Minnesota and its Department of Psychology for making their facilities available to us.

We also wish to acknowledge the many people whose assistance and participation were essential contributions to the study. Starke R. Hathaway, Elio D. Monachesi, and John E. Anderson gave us the benefit of their experience. Lawrence A. Young helped in the original formulation of the research design. Our interviewers, upon whom we depended for much of the data in this study, were: Douglas Freed, Arthur Gallese, Ronald Johnson, A. M. Marchionne, and Jay Willet.

The following directors and department heads of community agencies in the Twin Cities gave us their full cooperation in our search of agency files and many contributed much time to our study: Don Wilson of Minneapolis General Hospital Social Service, Harold B. Hanson of the Washburn Clinic, C. Wilson Anderson of the Minneapolis Family and Children's Service, Callman Rawley of the Jewish Family Service, Luthard O. Gjerde of Lutheran Welfare, T. F. Meagher of Catholic Welfare, Arnold Gruber of Hennepin County Welfare, H. B. Jensen of Minneapolis Welfare, Paul Keve and Donald McFarland of Hennepin County Court Services, Donald Winslow and Clifford Bailey of the Minneapolis Police, John Dothit of the Bureau of Criminal Apprehension, and Virginia R. Hathaway of the Minneapolis Board of Education Child Study Department.

Research Assistants on the project were Joan Drues, Claire Vedensky, Cheng-Ping Chú, Janis Klodt, and Alice Jensen. Other clerical staff members include: Constance Bowman, Jerome Downs, Virginia McHugh, and Mary Pearson. For most competent direction of the clerical staff we are grateful to Elizabeth Witt. The laborious secretarial tasks were ably handled by Kathryn Murray, Barbara McComes, Janis Pallister, Marcia Rutchick, and Louise Duus.

To all of them we wish to express our debt and appreciation.

Finally, and most importantly, we wish especially to thank the scores of individuals and their families who represent the samples reported here. They gave much of their time and their confidence to us, and we are most grateful.

Another way of defining delinquency designates an act delinquent which violates *any law.* This broad view is proposed by those who emphasize severity as important in a definition of "real" delinquency. The petty violations, so common as to be considered part of normal behavior, should not count.

In jurisprudence, the M'Naughten Rule—the interpretation of insanity based on the inability of the defendant to distinguish right from wrong—limits the interpretation of delinquency. More recently the defense of "irresistible impulse" has developed. The current, more liberal rulings allow broader psychiatric determinism. The definition of true delinquency is limited by the idea that in some cases the act is symptomatic of a psychiatric condition. From this position the definition would be restricted to include only those individuals who, in full cognizance of the law, their own situations, and the possible consequences, choose to commit an illegal act.

All of these definitions, and perhaps others as well, have merit. However, it seemed to the present investigators that the most defensible, inclusive, and objective position for scientific study was one in which the population to be included under the rubric *delinquent* was the one for which there was a known police or court record for apprehension and conviction. Within this group, on the double bases of frequency and severity of offense as legally defined, one can order the severity of the problem for the individual cases. It is possible to obtain an exhaustive sample of those individuals in any community in which records are kept, who have come to the attention of the police because of their troublesome behavior. These can be ranked with highly reliable objectivity. It means, of course, that one chooses not to attempt to study those "who were just as bad but did not get caught." While it is impossible to know the extent of that group, the opinion of the officials in the city from which the present sample comes is that the number of such delinquents is very small and they give the community little grief.

THE HATHAWAY-MONACHESI STUDY

The Hathaway and Monachesi (1953, 1957) study of Minneapolis ninth graders was begun in 1948. The population of *S*s for the study was the entire ninth-grade class of the public schools of Minneapolis during the academic year 1947-1948. Almost all class members, approximately 4000 students, were tested using the Minnesota Multiphasic Personality Inventory (MMPI).

After testing, a search was made, at two different times, to determine the delinquency patterns of the sample members; and, at the second time, a number of other social adjustment factors were examined.

Follow-up Procedures

The first rating of delinquency was made in 1950. At that time, the names of all cases were checked against the official records including the Minneapolis Police Force, the Hennepin County Sheriff's Office, the Crime Prevention Bureau, the Bureau of Identification, the Hennepin County Juvenile Court, the Court Services Probation Office, and the Municipal Court. The crimes found associated with the sample members were recorded and the date noted. A trained probation officer rated the individual records according to the four-point scale given below.

Rating 1. In this classification names were recorded in police records for at least one minor difficulty such as traffic contact (for example, overtime parking); or for being picked up but with involvement poorly established; or where the individual was contributing in such a minor way that it would not justify classification into one of the following groups.

Rating 2. The youngsters placed in this class had committed minor offenses such as destruction of property (especially when this was connected with play activities), drinking, one or more traffic offenses (escapades involving speeding, driving without a license, and/or going at high speed through a stop light or sign), curfew violation, and immoral conduct. The misbehavior was relatively nondelinquent in comparison to that of the other two categories. Nevertheless, these children as a group were considered to have demonstrated disquieting evidence of undesirable conduct.

Rating 3. This classification involved the commission of one serious offense such as auto theft, grand larceny, or gross immorality, or more than one less serious offense such as petty larceny, immoral conduct, assault, disorderly conduct, malicious destruction of property, shoplifting, flagrant curfew violations, truancy, and incorrigibility. The youngsters placed in this class were, therefore, not clearly established as delinquent, but nevertheless, they were showing behavior that needed more than casual explanation.

Rating 4. This level of misconduct was used to denote those who committed repeated offenses such as auto theft, burglary, grand larceny, hold-up with a gun, and gross immoral conduct (girls), accompanied by less serious offenses. All youngsters were placed in this category who were considered to have demonstrated a well-established delinquent pattern.

The second Hathaway-Monachesi follow-up study was completed during 1952, four years after testing. At that time, members of the study were approximately 19 years old. The follow-up procedures were the same as those used in 1950, but, in addition, the adult police files were surveyed. The second rating was of offenses occurring within the two-year period between 1950 and 1952. The same schedule of four points was used.

Sums of the ratings were obtained for the two periods in which cases were followed. In addition to the rating, it was noted whether the offense(s) occurred before testing, after testing, or both before and after the 1947-1948 school testing.

In addition to the delinquency follow-up in 1952, ratings of the *S*'s general adjustment were made by a social worker who called on the *S*s' families.[2]

MMPI Code and Delinquency Rate

Hathaway and Monachesi report that certain personality configurations appear to influence the rate of delinquency where they occur. After an extensive analysis of high-point code distribution, they have suggested that there exist two major dispositions relating MMPI code to delinquency. *Excitors* are codes that are associated with elevated rates of delinquency. The excitor effect is associated with elevation on the three scales: Psychopathic Deviate (*Pd* or 4), Schizophrenia (*Sc* or 8), and Hypomania (*Ma* or 9). *Suppressors* are codes that are associated with reduced rates of delinquency. Suppression is associated with three scales: Depression (*D* or 2), Masculinity-Femininity (*Mf* or 5), and Social Introversion (*Si* or 0).

A demonstration of classification based upon these dispositions can be seen in Table 1 where the data from the male population in the study are presented. The 2000 male cases were analyzed, first by trichotomizing the cases into a class of dubious delinquents who had combined rating through 1952 of less than 2, or were delinquent before testing, the remaining delinquent, and finally the nondelinquent cases. Within these three groups, all code cases were classified as to the two high-point code components.

TABLE 1

FREQUENCY OF CERTAIN CODE CLASSIFICATIONS AND DELINQUENCY LEVELS AMONG THE BOYS TESTED BY HATHAWAY AND MONACHESI

Code Classification	ND	Delinquency Status		
		Minor D and Before Testing	D After or Continued Through Testing	Total
Indeterminate	62	22	22	106
No High Point	31	10	4	45
Invalid	62	41	45	148
025	368.5	84.5	59.5	512.5
	(403)		(72)	
489	280.5	144.5	128.5	553.5
	(289)		(133)	
Other Codes	354	122	117	593
				1958

Note.—Parenthetical numbers are the number of cases in these classifications when the underlining in the code is disregarded. Underlined codes indicate that the scale scores from which the code is written are equal or within one point. Cases containing underlined codes generally are split, half a count in each of the code numbers that are underlined. The underline was disregarded in this study since groups were obtained by sorting IBM cards. Slightly different results would have been obtained from an exact code count.

Using the concepts of excitor (E), suppressor (S), and neutral (N) scales, an individual might have two excitor scales elevated (EE), a combination of excitor and suppressor scales (ES), or excitor and neutral (EN), or any other combination (viz., SE, SS, and SN, or NE, NS, and NN), giving nine combinations altogether as possibilities. Table 2 gives the concentration of delinquency for each combination and the frequency of delinquency subsequent to the ninth-grade testing.

These data indicate that when suppressors appear, they dominate the code, regardless of the character of the other code member. Thus the combination of two excitors gives a delinquency rate of 33%, while the combination of excitor with suppressor is about as low (or lower) than the 17% figure for combinations of two suppressors. Thus, any combination which includes a suppressor as one of the two deviant scales is associated with a tendency for the delinquency rate to be around the minimum found, regardless of the nature of the other high point scale.

THE PURPOSE OF THE PRESENT INVESTIGATION

While excitor codes together appear to be associated in greater than average degree with delinquency as a behavioral outcome, there were 67% of these individuals who did not become delinquent within the four years of the follow-up studies. On the other side, while the rate of delinquency among the individuals with elevated suppressor scales is low, there were 17% who did become delinquent. These findings lead to the questions: How do the individuals within these groups differ? What factors might be discovered in the environment, background, or personality development of some individuals who have what appears to be a delinquency-prone personality which leads some of them in fact to become delinquent, while others, presumably equally liable, according to the personality testing, do not act in antisocial ways? What factors in the lives of those with character structure assumed to be antithetical to delinquency are sufficiently influential to produce delinquent behavior despite such a personality make-up?

---

[2] Copies of research instruments designed for this study have been deposited with the American Documentation Institute. Order Document No. 6079 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $3.75 for microfilm or $2.00 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Since the records of Hathaway and Monachesi and the cooperation of the Minneapolis agencies concerned with these social problems were available, it was possible to set up a design which included samples of all individuals in the ninth-grade study who were known to have committed offenses. It was decided to study only the men, since the variety, severity, and frequency of delinquent behavior is much greater in boys than in girls. It is the male offenders who constitute by far the greater problem for the community. This left an original population of 1958 young men, then about 23 years old, all tested by Hathaway and Monachesi in 1947-1948.

Procedure

Subjects

Table 2 shows the cases classified into the excitor and suppressor groupings mentioned above. The differences between the numbers in parentheses and the remainder of the table is due to differing methods of classification. The numbers in parentheses give the number of individuals in the populations fitting the category in question when sorting cases according to code numbers and ignoring equal or nearly equal scales which had been split between code classes in the original code count.

Four groups were selected and matched for either code or delinquency status. These groups have been chosen from cases committing offenses after testing and who had a combined delinquency rating exceeding 1. Excluded were cases who were rated 2 because of two different traffic offenses, occurring in either or both of the follow-up periods. The two code classes were selected from the pure excitor codes or from codes containing a suppressor in the high-point code. Thus, by definition, the four groups selected for study were:

Group 1: Nondelinquent cases with an 0, 2, or 5 (henceforth 025) as one of the two high-point characteristics.

Group 2: Delinquent cases with an 0, 2, or 5, (henceforth 025) as one of the two high-point code characteristics. (This includes only those delinquent after testing and omits minor delinquents.)

Group 3: Nondelinquent cases with a 4, 8, or 9 (henceforth 489) as both of the high-point code characteristics.

Group 4: Delinquent cases with a 4, 8, or 9 (henceforth 489) as both of the high-point code characteristics. (This includes only those delinquent after testing and omits minor delinquents.)

Group Characteristics. Using social worker ratings made by Hathaway and Monachesi in 1952 (see Footnote 2), the four groups in the present study were compared. The significant differences associated with the groups on the 12 ratings are reported for the code, the delinquency and within

TABLE 2

Boys Studied by Hathaway and Monachesi in 1947-1948 Having Excitor (4, 8, 9), Suppressor (0, 2, 5), and Neutral (1, 3, 6, 7) MMPI Codes and the Percentage of Delinquent Cases[a]

Delinquency Status	EE	EN	NE	ES	NN	SE	NS	SN	SS
No. Nondelinquent	274.5	165	81.5	118	19	74	19	37	42.5
No. Delinquent	138.5	57.5	25	22	9.5	11.5	7.5	5.5	8.5
% Delinquent in column	33	26	23	16	—	13	28	13	17

Note.—Omitting cases with delinquency ratings of less than two and cases that were delinquent before testing but not after testing.

[a] Code characteristics: Suppressor (S), Excitor (E), and Neutral (N).

code group (delinquency) comparisons. Results showing stability at the .05 level of significance are reported. (*a*) A dichotomy of *normal* vs. *broken home* shows Group 2 to contain a significantly larger number of broken homes than the other three groups. (*b*) The measure of *family mobility before testing* shows the delinquent groups (Groups 2 and 4) to have moved more often than the nondelinquent groups (Groups 1 and 3). (*c*) Two education measures, years of high school completed and education after high school, show both delinquency and code type to be differentiating variables. The nondelinquent groups have the greatest success on both measures. It is clear, however, that the individual 025 groups exceed their 489 mates. (Group 1 is better educated than Group 3, and Group 2 is better educated than Group 4.) (*d*) The measures of the *neighborhood rent level* show that the delinquent groups come from lower rent level areas. (*e*) The rating of the *S*s' *work records* shows that the nondelinquents have better work records than the delinquents. (*f*) Ratings of constructiveness of *home, community,* and *leisure time activities* reflect both MMPI code and delinquency status differences. Again, as in the case of the measures of education, good ratings on home, community, and leisure time activities are positively related to nondelinquency. The 025 groups also exceed the 489 groups on these measures. In descending order from good to poor constructive use of time, the groups are 1, 2, 3, 4. (*g*) Group 1 registers more *physical illness* than the other groups. This difference is slight but reliable. (*h*) Group 3 registers less *personal maladjustment* as rated by the interviewer than any of the other three groups. (*i*) Rather incomplete intelligence testing by the schools seems to reflect more *low intelligence* people in Group 4.

Sampling Procedures

From three of the four groups, samples were drawn with the intention of obtaining representative samples of approximately 75 cases. The whole of the appropriate population was accepted for Group 2 since it contained only 72 individuals. The three samples for Groups 1, 3, and 4 were drawn randomly from the other parent populations.

Upon the incorporation of further information from the 1956 follow-up (discussed below), some cases seemed no longer to belong with the original population and had, therefore, to be excluded from the samples after they were randomly selected. The following gives an account of why, when 75 cases were selected, the *N*s in these three samples were not 75.

Because of clerical error, the original selection for Sample 1 contained an extra case, making $N = 76$; similarly only 74 cases were picked for Sample 4.

Sample 1 (nondelinquent 025's), drawn from the population of 409 cases, originally consisted of 76 cases. After an extensive check of police and court records, six cases were found to be delinquent. Two of these had been missed in the 1952 follow-up in the Hathaway-Monachesi study. The remaining four delinquents had been in trouble between 1952 and 1956. The two who were delinquent before 1952 were excluded as were three of the remaining four. One was retained since his summed ratings were not greater than one. The resulting sample contained 71 cases.

Sample 3, the nondelinquent 489 sample, was drawn from Group 3, which contained 289 cases. From this population, 75 cases were drawn. Here again, two cases with records of delinquency that had been missed in the earlier work were found. These cases were rated greater than one and thus were excluded. Four other cases showed records of delinquency after 1952. Using the same criterion as with Sample 1, two cases were excluded due to the severity of the records. The resulting Sample 3 contains 71 cases.

Sample 2 (delinquent 025's) contained only 72 cases. In the recent extensive record check one of these was found to be nondelinquent. The sample (or, more accurately, the population) thus contained 71 cases.

Sample 4, the delinquent 489 sample, was drawn from a population of 132 cases. Seventy-four cases were drawn, not 75. One of these cases was not delinquent on further check of the records. The sample thus contained 73 cases.

The cases in each sample, an estimate of the number of cases in error, and the true population *N*s appear in Table 3. These estimates are made from the proportions of cases in error in the samples.

Sampling Adequacy. The variables originated in the Hathaway-Monachesi study were again used to check the randomness of the samples. Using each of the ratings mentioned above in the description of the groups, the sample and the remaining members of the parent groups were compared. The three delinquency measures are only mentioned in connection with Sample 4 (since all of the appropriate population was selected for Sample 2). In general, sampling appears to be adequate, for there were only two minor differences between the samples and the remainder of the groups. Sample 1 showed a disproportionately large number of individuals rated as making a "good" adjustment in *community affairs*. Sample 4 seemed to be somewhat too well filled with cases getting a middle range rating on *home adjustment* at the expense of cases with a "good" rating. Both of these differences are reliable at the .01 level. Neither seems to offer much suggestion that the samples are enough different from the populations to make them additionally suspect.

Sample Characteristics

MMPI. A comparison was made between the samples with like MMPI code types. The mean MMPI profiles for each of the four samples are

TABLE 3

CASE CLASSIFICATION

	Not Delinquent	Delinquent
	Sample 1 = (dl)	Sample 2 = (Dl)
MMPI code type 0, 2, 5 Delinquency Suppressor	Boys with nondelinquent personality disposition and who were in fact not delinquent. (Test-*Hit*) Population N = 409 (398) Sample N = 71	Boys with nondelinquent personality disposition but who in fact became delinquent. (Test-*Miss*) Population N = 72 (71) Sample N = 71
	Sample 3 = (dE)	Sample 4 = (DE)
MMPI code type 4, 8, 9 Delinquency Excitor	Boys with delinquency-prone-personality who never became delinquent. (Test-*Miss*) Population N = 289 (281) Sample N = 71	Boys with delinquency-prone personality who did in fact become delinquent. (Test-*Hit*) Population N = 133 (129) Sample N = 73

presented in Figs. 1 and 2. Comparisons were made between the cases in Samples 1 and 2 (025's) and between Samples 3 and 4 (489's); that is, between those samples with like code types. (Selection procedures have already insured that other differences occur.) These comparisons suggested reliable differences between Samples 1 and 2 on one scale, *Pd*, and also between Samples 3 and 4 on the *Pd* scale. In each comparison the delinquent sample mean exceeded the nondelinquent sample mean.

Delinquency Rating. In order to classify as accurately as possible all cases with regard to their delinquency status, a third follow-up was undertaken in 1956. For the sample cases, the records were checked again in the process of completing this study. The earlier criminal records of all cases were verified or corrected, and a new delinquency rating, using the same four-point scale, for the period between 1952 and mid-1956 was given all cases in the study.

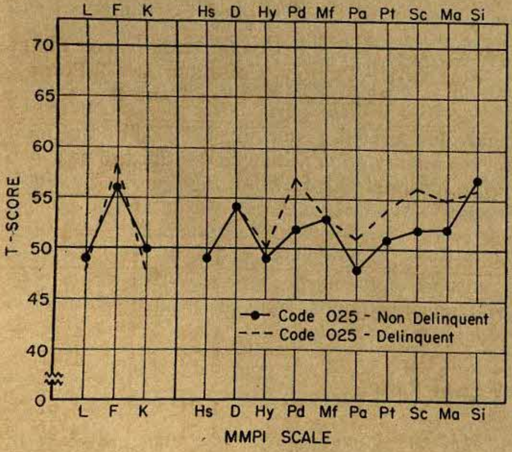


FIG. 1. Mean profiles for Samples 1 and 2 from the 1948 MMPI testing.

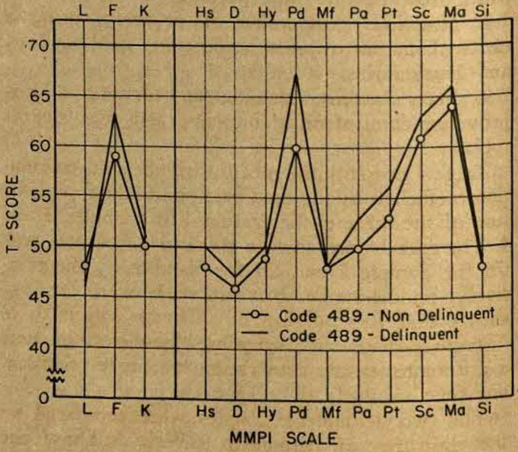


FIG. 2. Mean profiles for Samples 3 and 4 from the 1948 MMPI testing.

The density of delinquency in both samples decreased with time. The percentage of cases delinquent in each of the three periods was 80 (1950), 63 (1952), and 21 (1956) for Sample 2; for Sample 4, these percentages were 86, 57, and 22. The mean ratings also decreased with pronounced steps, reflecting, of course, the number of zero ratings entering into the sums.

Since there were three follow-up periods, any individual's delinquency rating could be summed over the entire period, giving a possible range of 0–12.

The three ratings given each individual within the delinquent samples were analyzed using Fisher's intra-class correlation. The resulting F ratio was insignificant. This was thought to invalidate a procedure using the summed rating as a measure of severity. Thus, a final judgmental measure of delinquency was derived from the offense records covering the entire period to mid-1956. The judgment, then, includes all offenses attributed to the individual. The junior author and Lawrence A. Young, Research Director for the Hathaway-Monachesi study, sorted freely, into an unrestricted distribution with a range of nine points, attending to the psychological significance of the records by considering: (*a*) the severity of the offenses, (*b*) repetitiousness of the record, (*c*) the chronicity of the record (as contrasted to being temporarily circumscribed), and (*d*) the age at the occurrence of the offenses. The correlation between these two judges was .95. The cases in disagreement were examined, and a compromise rating was made by one of the judges after attempting to understand the nature of the difference.

One case typical of each of the odd-numbered levels is reported to illustrate the scale.

1. One incidence of purchasing and drinking liquor when underage.
3. Breaking school windows and driving without a license.
5. Immoral conduct, receiving stolen bicycle parts, prowling cars, and speeding.
7. Malicious destruction of property, using a car without the owner's permission, stealing cars, and drunkenness.
9. Petty larceny, stealing and driving a car, prowling cars, stealing bicycles, and first degree robbery.

Table 4 presents the mean, variance, and correlation coefficient of the final nine-point rating and the sum of the three earlier ratings for Samples 2 and 4. The correlation between these two measures was .76 for Sample 2 and .79 for Sample 4. The relatively middle-range correlation between ratings shows that the judging of offenses within three discrete periods and summing these ratings nets a different measure than rating the entire record at one time.

The two delinquent samples were compared on the measures of delinquent severity. These are the ratings made in 1950, 1952, and 1956. The ratings were summed and the new nine-point ratings were also compared. The samples do differ on the mean 1952 rating, when the cases getting zero ratings are omitted ($P < .01$). Here Sample 4 was rated as more severely delinquent than Sample 2 (mean ratings 2.70 vs. 2.07).

Types of Crimes. Table 5 presents the frequency of each of 25 types of offenses for the two delinquent samples prior to 1953. The figures in Table 5 are roughly comparable. (None of the frequencies in Group 2 is large enough to be adjusted upward as if the base were made 73 cases.) Table 5 enumerates all known crimes for the individuals in these two samples. Of these, 204 were committed by people with 489 profiles and 194 were committed by people with 025 codes. The adjusted values for these are 204 and 197. The breakdown into types of crimes suggests that crimes of profit are slightly more common in the 489 group. The adjusted values are 72 for Sample 2 as against 77 for Sample 4. The adjusted record for the two samples on conduct offenses is 53 vs. 61 for Samples 2 and 4. Among the purely traffic offenses adjusted values are 72 vs. 66 for Samples 2 and 4, the only reversal.

FURTHER TECHNIQUES OF CASE STUDY

Social Agency Contact by the Families of Ss

The source from which the data were derived was a survey of social agency records pertaining to families of the *S*s. Certain family and community interactions were found to be related to the personality and delinquency outcomes in *S*s. These data, therefore, included contacts by family members other than the *S*s themselves.

While the nature of the agency contact has, of course, changed over the last 25 years, all agencies have kept records of their contacts with families in the community.

TABLE 4

MEAN, VARIANCE, AND CORRELATION (*r*) FOR THREE FOLLOW-UP PERIODS COMBINED AND 9-POINT RATING ON SAMPLES 2 AND 4

Rating and Parameter	Sample 2	Sample 4
Combined Rating		
$\bar{X}$	3.89	4.33
Var	2.76	4.47
9-Point Rating		
$\bar{X}$	4.21	4.62
Var	3.71	4.49
Correlation	.76	.79

The agencies surveyed include the municipal relief agency, the county relief agency, three religious welfare agencies, one Community Chest family and children's service, the state and city hospitals, the Child Study Department of the Minneapolis Public Schools, the only community psychiatric clinic for children, and the county Court Services.

TABLE 5

Offenses Committed by Samples 2 and 4 through 1956 Follow-Up

Type of Crime	Sample 2	Sample 4
Crimes of profit		
Auto theft	14	21
Burglary	10	15
Grand larceny	4	6
Petty larceny	14	20
Shop lifting	3	1
Destruction of property	9	11
Receiving stolen goods	2	3
Drug traffic	2	0
Contact only	13	0
Totals	71	77
Corrected total[a]	72	
Conduct offenses		
Truancy	1	2
Curfew violation	2	1
Absenting	0	0
Disorderly conduct	16	18
Immoral conduct	7	12
Assault	2	11
Probation violation	0	13
Liquor	16	19
Incorrigibility	4	3
Other	4	2
Totals	52	61
Corrected total[a]	53	
Traffic offenses		
Drunken driving	11	4
Careless driving	9	9
Faulty operatives	4	4
Parking tickets	8	11
Driving without license	11	18
Speeding	15	10
Failure to stop at stop signs	7	7
Other	6	3
Total	71	66
Corrected total[a]	72	
Total for all offenses	194	204
Corrected total[a]	197	

[a] Corrections are adjustment for the difference between the 71 cases in Sample 2 and the 73 cases in Sample 4.

The case records of all boys and their families which were identified in the admission files of the 11 agencies were completely reviewed for the data which seemed psychologically important. The investigation was carried out by three graduates of the University of Minnesota, Department of Psychology. This procedure was followed except at the municipal hospital where a social service summary was read instead of the chart, and at the county welfare agency where records were reviewed by the case work auditor. Data applying to an *S* were, of course, set aside since the survey was aimed only at the family.

The data collected in note form, as described above, were found to include 42 fairly common but discrete items. These 42 items were then grouped into seven more general categories (see footnote 2). All events were counted once for each year in which they occurred. The annual totals were grouped into the following seven age periods: (*a*) before the *S*'s birth; (*b*) birth–2 years (infancy); (*c*) 3–5 years (early childhood); (*d*) 6–12 years (childhood); (*e*) 13–15 (early adolescence); (*f*) 16–21 (adolescence); (*g*) 21 years and on (young adult).

After combining all the information gathered from the agencies, the raters summarized the points within each category and also within each age period.

Although each category is self-descriptive, a few comments may serve to clarify the meaning of each. *Family Disruption Due to Disease* was made up of items that simply describe major events in a family caused by disease. These are presumed to be destructive of family morale. *Poverty or Economic Need* included complaints of financial need brought to agencies for solution. *Dissocial Behavior in the Family* was a conglomeration of family actions that reflect delinquency on the part of family members other than the *S*s. *Psychiatry for the Family* reflected factual instances of agency contact which included psychiatric diagnosis or treatment. Hearsay or conclusions by the agency staff were recorded in another category, *Minor Psychological Problems. Marital Disruption* was a record of formal interruptions such as divorce. These behaviors were quite unlike the data in the category of *Dissocial Behavior in the Family. Inadequate Parent-Child Relationship* included a number of economic or physical mistreatments of children by their parents. The last category, *Minor Psychological Problems,* was really a record of counseling case work aimed at adjustment problems and recorded family conflicts.

Evaluation by Parents

The cooperation of the parents of the *S*s was enlisted to provide another source of data. In the present study, objective recording of information and standard coverage of case material were considered essential. Earlier work by Briggs (1955) dealt with the use of a self-administered History

Record (M-B History Record) designed to be completed by an informant. This technique provided the means of recording and coverage specified and has the additional advantage of requiring relatively little time (and no professional time). Briggs' work suggests that such a technique can have more complete coverage than a history obtained from an interview. Some evidence also exists to show that common history concepts can be defined in terms of item sets, as they appear in a standard objective history (Briggs, in press).

The History Record used in this investigation contained 138 items referring to the "subject." These items are designed to contain an alternative answer for any person in the present population (see Footnote 2). Families were called originally to make appointments for an interview with the *S*s. After the family had been located, the mother of the *S* was called by telephone to enlist her cooperation. Copies of the History Record and stamped return envelopes were sent to each informant. Post cards and telephone calls were used as reminders.

The Clinical Interview

The method which it was hoped might provide the most revealing information was the interview. The interview is central to the clinical approach. The *S*s' histories, past and present attitudes, and personality are most easily available through interview. The usual clinical interview is open, leading into the areas most interesting to the clinician. For research purposes, there must be standard case coverage and objective recording. Four areas were selected for concentrated investigation: first, the history, especially in the years following the ninth-grade testing; second, family attitudes as demonstrated during adolescence; third, the general adjustment of the *S* in social relations, in occupational and in educational areas; and fourth, the present personality and dynamics of the *S*s.

The entire interview was structured for recording purposes. All areas were itemized, and whenever possible, ratings or check lists were provided for recording of data. The answers to open-end questions were briefly recorded. These were classified after the interviews were completed.

A great number of avenues were used to find the *S*s. First contact with the *S* or his family was made by telephone. It was explained to the *S* that almost everybody who is sick or in trouble is asked to give a kind of history of his life, that nobody had ever taken histories from normal people, and that this is what the psychology department of the university now wished to do. The interviewers were persistent in their attempts to obtain an interview.

All of the interviewers were male. Of the five individuals conducting more than one interview, one had a PhD in psychology, three held masters' degrees in psychology, and one had a bachelor's degree with four years' experience as a probation officer.

The interview consisted of 157 items. Of these, 11 were ratings, 135 were lists, 11 were open-ended items having unspecified answers, and the remainder were identifying information (see Footnote 2). A number of items overlap or are identical with the items in the M-B History Record, which was described above. There were 23 items exactly duplicated in the two methods and another 15 items with comparable answer choices.

Besides the historical and attitudinal material, other items were included to test specific hypotheses. First, ten items were selected from the Gluecks' (1950) study. These items were included to compare their power in this study with the material reported by the Gluecks, who found them highly differentiating predictors of delinquency.

Secondly, an item called "*Pd* signs" included a count of the number of visible characteristics that mark identification with the stereotype of the "wilder" subcultures of society. Since most of the characteristics here have been interpreted by society, the item provides an interesting test of the validity of such judgments (or prejudices). Some of these characteristics are wearing a leather jacket, long sideburns, "duck tail" haircut, and owning a "hot rod" or modified automobile.

Two other items were constructed to estimate factors that might be explanatory of delinquency. One of these, Opportunities to Act Out, is the sum of a number of items believed to be associated with making it easier to get into trouble (e.g., free time, absence of supervision, neighborhood delinquency, etc.). This item was a hypothetical cluster of "causes of delinquency" gathered under the single concept of opportunity. Another similar item, Changes in Life Adjustment after the 9th Grade, included a count of the number of posttesting changes in history that might encourage delinquency. This item was specifically designed to reflect happenings that might have influenced cases with 025 codes (Group 2) to become delinquent. Some of these changes are: school changed, identification figures changed, and moved to a different neighborhood.

The Q Sort. In order to obtain a quantifiable index of the interviewers' appraisal of the *S*s, each interviewer sorted a *Q* deck just following each meeting with an *S*. The deck of items adopted for the study was developed by Block (1954) for the Institute of Personality Assessment and Research. The items checked in the formal interview did not include the area of present personality, since the interviewer rated these by his sort. The interviewers were instructed to conduct their interviews with the *S*s in such a way as to be able to sort the *Q* deck adequately.

RESULTS

The findings of the investigation are reported, first taking the items individually. Following this, the results for the complex clusters of items and the overlapping data are reported.

Social Agency Contact by the Families of *S*s

In recording agency contact by family members, data on the *S*s' parents and siblings alone were recorded. These contacts tell something of the nature of the family environment in which the *S*s lived.

Four levels of the agency contact are considered below: the fact of contact by the family as opposed to no agency contact, the total number of items which are true (counted as points) per sample, the number of such points for each item in each sample, and the number of points in each category associated with the samples. (The more basic data involving agency contact appear in Table 7.)

Contact vs. No Contact

The number of families who had agency contact by each age period was analyzed. Tests of independence at each age level failed to suggest that there was statistical significance attached to the difference in frequency of agency contact for the two code groups (Samples 1 plus 2 vs. Samples 3 plus 4). When the same data were combined into totals for the delinquent and nondelinquent groups (Samples 2 plus 4 vs. Samples 1 plus 3), reliable differences were found at all age levels above age 2 ($P < .01$ for each age period). In essence, then, the subsequent delinquency or nondelinquency, but not the personality characteristics of the boys studied, were anticipated by a family record with social agencies. This anticipation is just a suggestion, however, and is not useful in itself as a practical predictive sign.

Although there was no reliable relationship between MMPI code type and occurrence of agency contact, the important relationship between MMPI code and subsequent delinquency has already been emphasized. It was this relationship which was the major finding in the Hathaway-Monachesi study and an original basis for group selection in the present study. Taking both of these factors into account, Multiphasic code and agency contact, it is possible to state the rate of delinquency estimated from their combined effects. These values were calculated from population estimates in Table 4 and the data in Table 8. Of the group of boys who had 489 codes and whose families had contact with a social agency, 42% became delinquent. Of the group of 489 boys without such agency contact in their families only 23% were delinquent. Of boys with 025 codes plus families with agency contact, 22% became delinquent, while of boys with similar personality test patterns, but from families with no agency contact, only 11% were delinquent.

Total Points Per Sample

Methodological Considerations. Before examining the content of the agency contacts, it should be kept in mind that arbitrary decisions determined two methodological issues. First, each rater was instructed to give one point per item for each year in which the item occurred. Second, the content of items was used as an a priori guide in their grouping into the seven categories used. No tests were made of item homogeneity or of weighting for scaling, and it is possible that such procedures would refine the results reported here.

Furthermore, the categories differed greatly in composition. Certain items concerned occurrences that were repeatable many times (e.g., chronic aid); others may occur only once or a few times in a lifetime (e.g., death of a parent or divorce). Within categories of predominantly low rate items, small totals were the rule regardless of the strength of the relationship between the items and delinquency or personality pattern.

Finally, it is already known that the groups are unequally represented among the cases who had a history of agency contact in their families. There was a total of 86 cases among the 145 delinquents studied who had families with social agency records; 56 nondelinquent cases came from such families. Category sums were limited, therefore, by the number of cases who could contribute.

Three comparisons were made among the samples to test for personality differences, delinquency differences, and for severity of delinquency within the delinquent groups. In determining severity, the 9-point rating made in 1956 covering the whole record was dichotomized, and those cases with ratings of 9 through 4 were called *mild,* while cases rated 5 through 9 were called *severe.*

Total Points Per Sample. In terms of total points for all of the categories of social agency contact, the families of Sample 1 obtained 132 points; for Sample 2, the figure was 361; for Sample 3, 178; and for Sample 4, 434. After these totals were adjusted for the inequalities in representation of families concerned, important differences remained. The results are present in Table 6.

Code Comparisons. Chi square tests failed to suggest interdependence in the distribution of points per case either in delinquent (Sample 2 vs. Sample 4), nondelinquent (Sample 1 vs. Sample 3), or total code (Samples 1 plus 2 vs. Samples 3 plus 4) comparisons. This was congruent with the code group comparisons for cases with agency contact.

Delinquency Comparisons. A relationship was found between the (per case) totals and delinquency or nondelinquency within each code group and also when taken combined (see Table 6). In essence, the delinquents' families accumulated more than twice as many points.

The severe delinquents' total significantly exceeds the mild delinquents'. This trend was not strong enough to be registered within either sample taken individually. These results are also presented in Table 6.

TABLE 6

SAMPLE DIFFERENCES ON POINTS RECORDED FROM AGENCIES

(Expected frequencies calculated per *S* with a family having a history of agency contact)

Samples or Groups Compared	Significance	Group Having Larger Sum
D vs. ND	.01	(2 and 4)
1 vs. 2	.01	(2)
3 vs. 4	.01	(4)
Mild vs. severe	.01	(Severe)

Items Differentiating Between Samples. The frequency of each of the items for each of the four samples is presented in Table 7. The data were analyzed for two periods: before the *S*s' 13th birthday and from age 13 to the present time (1956). This division roughly approximated the division between adolescence and childhood, and was before the age of the ninth grade testing for all cases.

Code Differences. Samples 1 and 2 each contained boys with 025 code types while Samples 3 and 4 had boys with 489 codes. Items which differentiated them may shed light on some of the family situations which are associated with the development of one kind of personality pattern in the children in the first two samples and another personality outcome in the other two samples.

Those items which differentiated between code types for either of the major age periods are reported in Table 8. In all instances of significant relationship, the larger totals were associated with the histories of boys with 489 codes. Comparisons were based upon the per case with agency contact paradigm. One item, child removal, showed this relationship within the first 13 years, and three items showed the same trend in the later period. These were: chronic illness on the part of a sibling, crime by a family member, and social case work for the family. For the total period, child removal and social case work indicated a relationship between point totals and 489 code types. Comparisons of delinquents and of nondelinquents with differing codes (Sample 1 vs. Sample 3 or Sample 2 vs. Sample 3) were omitted because the relatively small totals associated with the individual items did not permit comparison.

Delinquency Differences. Samples 2 and 4 contained boys with known records of delinquency. Samples 1 and 3 were composed of boys who were nondelinquent. Items differentiating delinquents and nondelinquents may give some indication of the family stress which contributed to the development of delinquency in the children

TABLE 7

FREQUENCY OF OCCURRENCE OF EACH HISTORY ITEM IN THE FOUR SAMPLES FOR THE PERIOD BEFORE AND AFTER AGE 13

Item	Before 13 Years				After 13 Years			
	Sample				Sample			
	1	2	3	4	1	2	3	4
Mother dies		4			1	6		3
Father dies	2		1	3	2	2	3	1
Parent has chronic illness	10	19	10	30	7	13	8	13
Sibling dies	1	6	1	5		6	2	
Sibling has chronic illness	8	26	5	25	1			10
Multiple OB	2	3		5				
Chronic aid	24	82	35	84	19	27	22	18
Temporary aid	4	6	3	5		3		1
Excessive debts		1		2		1	1	2
Housing needs		2			1			3
Unmet financial need	3	5		5		1		1
Maternal abortion	1	4	2	6				
VD				1				
Illegitimate parentry	4	5	1	5	1			3
Desertion	1	5	2	2	1	5		3
Workhouse-Jail	1	10	4	6		3	1	2
Deleterious drinking		7	4	5				1
Crime, probation, problems with the law	5	25	8	21	2	14	8	29
Employment problem	1	3	2	6			1	3
Complaints (by others)	1	5	2	3	1	1	2	5
Mental retardation	1	5	2	4	1	2	2	8
Psychiatry contact		2		1				
Suicidal			1		2	1	1	2
Alcoholism			2	4	1		1	2
Undiagnosed mental illness			2	3		1		1
Divorce	2	2	1	1		2	1	6
Separation	1	2	1	3	2			1
Absence of parent		4	2	2		1	1	
Promiscuity	3	3	1	4	2	3	1	5
Neglect		1		2				
Abuse								
Non-support	1	6	4	5	4	4	3	5
Child removal		1		7		3	3	3
Social service case or counseling	2	3	2	10	1	4	3	19
Marital adjustment problems	1	3	1	3	1	1	1	2
Parent child conflict		1	1	2		3	1	2
Department of child study	5	11	9	11	7	14	13	18
Hospitalization, mental	1	2	6	3	2	1	6	5

TABLE 8

ITEMS MORE OFTEN ASSOCIATED WITH THE FAMILIES OF *S*s WITH 489 CODES THAN WITH 025 CODES

(Probability levels are based upon χ^2 or *z* distributions)

Item	Age Period		
	To 13	After 13	Total Period
Sibling has chronic illness		.01	
Crime by family member		.02	
Child removal	.05		.05
Social service case or counseling		.01	.01

or, on the other hand, patterns of social agency contact related to the development of social noncomformity on the part of the children.

Items that occur differentially within the histories of the delinquent and nondelinquent groups appear in Table 9. Six items were found more frequently in the family histories of delinquent boys in the period before the children were 13 years old. Three of these concerned death or health problems; chronic illness of a parent, death of a sibling, and chronic illness of a sibling. The other three items related to subsequent delinquency were: a record of crime by family members, the history of removal of a child from the home, and chronic aid.

After age 13, the following four items related positively to delinquency: the death of a mother, the chronic illness of a sibling, a crime by a family member, and social service counseling. It is unknown which instances occurred after acts of delinquency, although some certainly must have. Again, the individual sample comparisons were omitted because of the small frequencies involved.

Categories of Agency Contact Reflecting Differences

Sample totals were obtained for the seven categories of agency contact. These values were analyzed for differences that pertain to code, delinquency, and severity of delinquency.

Code Differences. Only one category, *Psychiatry for the Family,* registered a difference before age 13. This was between the nondelinquent Samples 1 and 3, in which the 489 sample had more psychiatric contacts. For the period after age 13, *Psychiatry for the Family, Dissocial Behavior in the Family,* and *Minor Psychological Problems* reflected a code group difference (Samples 1 plus 2 vs. Samples 3 plus 4) in the *S*s. These differences appeared to originate in the Sample 2 vs. Sample 4 comparison, with greater point totals in the families with boys having 489 code types. These differences are reported in Table 10.

Since some of the relationships noted above must have occurred after the testing of the boys, the interpretation of results is ambiguous. Nevertheless, these suggest that considerable maladjustment was present in other members of families producing boys with 489 personalities.

Delinquency Differences. The categories relating to delinquency appear in Table 11. Three categories related to delinquency within the 025 samples before the age of 13 years. These were *Family Disruption Due to Disease, Poverty or Economic Need,* and

TABLE 9

ITEMS MORE OFTEN ASSOCIATED WITH THE FAMILIES OF DELINQUENT THAN NONDELINQUENT *S*s

(Probability levels are based upon χ^2 or *z* distributions)

Item	Age Period		
	To 13	After 13	Total Period
Mother dies		.01	.01
Parent has chronic illness	.05		
Sibling dies	.05		
Sibling has chronic illness	.01	.05	.01
Chronic aid	.01		.01
Crime by family member	.01		.05
Child removal	.01		.05
Social service case or counseling		.01	.01

Dissocial Behavior in the Family. The first two of these categories were more characteristic of the 489 delinquent sample than the 489 nondelinquents before the 13th year. Both samples of delinquents exceed the nondelinquents (1 plus 3 vs. 2 plus 4) in the number of points classified under *Inadequate Parent-Child Relationship,* as well as in the three categories mentioned above.

In the second age period, past the 12th year, Sample 2 accumulated more points than Sample 1, and Sample 4 more than Sample 3, in the category *Dissocial Behavior in the Family.* Sample 4 had more points also in *Psychological Problems* than Sample 3.

Delinquent Severity. The difference between the family histories of the mild and the severe delinquents occurred mainly before the 13th year. Categories showing significant relationships appear in Table 12. Except for family disruption due either to illness or to divorce, all forms of family maladjustment are more frequent in the severely delinquent children's homes. In the later age period, disease and marital disruption appear to relate to severity, while poverty and inadequate parent-child patterns do not. The other categories, dissocial behavior and psychiatry for the family, retain a relationship throughout the two periods. These results suggest that the

TABLE 10

SOCIAL HISTORY CATEGORIES MORE OFTEN ASSOCIATED WITH THE FAMILIES OF THE 489 SAMPLES THAN THE 025 SAMPLES

Category	Age Period					
	To 13	After 13		Total Period		
	1 vs. 3	025 vs. 489	2 vs. 4	025 vs. 489	1 vs. 3	2 vs. 4
Dissocial behavior		.01	.05		.05	
Psychiatry for family	.01	.02	.02	.02	.02	.01
Minor psychological problems		.01	.02	.01		.02

TABLE 11

SOCIAL HISTORY CATEGORIES MORE OFTEN ASSOCIATED WITH THE DELINQUENT SAMPLES THAN THE NONDELINQUENT SAMPLES

Category	Age Period								
	To 13			After 13			Total Period		
	1 vs. 2	3 vs. 4	D vs. ND	1 vs. 2	3 vs. 4	D vs. ND	1 vs. 2	3 vs. 4	D vs. ND
Family disruption	.05	.01	.01				.01	.01	.01
Poverty or need	.01	.01	.01				.01		.01
Dissocial behavior	.01		.01	.05	.01	.01	.01	.01	.01
Inadequate parent-child relation			.02						
Minor psychological problems					.02	.01		.01	.01

TABLE 12

SOCIAL HISTORY CATEGORIES MORE OFTEN ASSOCIATED WITH THE FAMILIES OF THE SEVERE DELINQUENTS THAN THE MILD DELINQUENTS

Category	Age Period		
	To 13	After 13	Total Period
Family disruption		.05	.01
Poverty or need	.01		.01
Dissocial behavior	.01	.01	.01
Psychiatry for family	.05	.05	.01
Marital disruption		.01	.05
Inadequate parent-child relationship	.01		
Minor psychological problems	.01		.01

severely delinquent boy comes from a home that is more disturbed than the delinquent generally.

Time of Agency Contact

The total number of points per page for the *S*s' families was examined. Differences among age periods may reflect differences in the economic climate of the nation, which presumably affected all of the families. Differential outcomes among the groups at any given age period may be helpful in understanding the various family reactions to this general life stress. The average year of birth for *S*s in each sample was 1933, a time when the national economy was severely depressed.

The interaction between sample *N*s and age-period totals was explored for differences relating to personality, delinquency, and severity of delinquency among *S*s. The periods of greatest relationships for each of these show that the delinquents' families had more trouble in the period from before the *S*s' birth to the third year ($P < .02$), while the families of the nondelinquents had more trouble after the boys' 21st year. The *S*s with 025 codes had families who experienced their maximum period of trouble when the boys were from 2 to 5 years old ($P < .05$), while the families of boys with 489 codes had most trouble in the period when the *S*s were 16 to 21 years old ($P < .01$). No relationships occurred between severity and time that are not anticipated in the marginal totals.

Interrelationship of Categories. Finally, it is of interest to examine the interrelationships between the categories of social agency contact. The three categories of *Family Disruption Due to Disease, Poverty or Economic Need,* and *Dissocial Behavior in the Family* accounted for approximately three-fourths of the 1105 points accumulated. Each distribution of points for all individuals was dichotomized as nearly at its median as was possible. Phi coefficients for the 142 cases were calculated for these three pairs of categories. This procedure was followed for the totals up to the 13th year and for the totals over the entire period. These results appear in Table 13. Before the 13th year, each of the three phi coefficients was significantly greater than zero. The largest of these (.37) related *Poverty or Economic Need* with *Dissocial Behavior.* For the entire period only one correlation, that between *Family Disruption* and *Poverty,* was above a chance level (.33, $P = .01$).

These results present some interesting problems. First, one wonders why these values were highest in the earliest period

TABLE 13

MATRIX OF PHI COEFFICIENTS RELATING THREE CATEGORIES FOR 142 FAMILIES WITH AGENCY CONTACT

Category	Category		
	Poverty or Need	Dissocial Behavior	Dichotomized at:
Family disruption			
to age 13	.28**	.21*	$1 \leq 1$
Entire history	.33**	n.s.	$2 \leq 2$
Dissocial behavior			
to age 13	.37**	—	$1 \leq 1$
Entire history	n.s.	—	$1 \leq 1$

* $P < .05$
** $P < .01$

and dropped off in the later period. Possibly the economic condition of the nation makes the difference. In a depression, paying hospital bills, frequenting relief agencies, and committing crime may reflect the same family problems to a greater extent than when prosperity characterizes the national scene. A second hypothesis is that families with younger children may be put under a more uniform kind of pressure, and, consequently, the kinds of behavior that they exhibit are more uniform. Later, as the family matures, they become less similar.

In addition to intercorrelative relationships in the history, it is meaningful to examine relationships between the categories and the degree of delinquency. An aspect of this question has been answered since we know that the families of severe delinquents accumulated more points on many variables than the families of mild delinquents. The correlation coefficient for the categories discussed above failed to show a relationship between the variables involved and delinquency rating.

Summary

In this section, relationships between agency contact and delinquency, severity of delinquency, and MMPI code type were examined. From these results it is apparent that simple contact with a social agency by any member of a boy's family is reliably associated with a history of delinquency in the boy himself, if taken with knowledge of the code group in which the boy belongs.

Within the content recorded at social agencies, certain kinds of historical items occur frequently in the delinquent boys' families. These items relate to *family disruption, dissocial behavior,* and *psychological problems.* The number of points associated with the delinquent samples is most differential in the period before birth.

Among the boys who were delinquent, the severity of delinquency was related to three types of family disturbances: *poverty and economic need, dissocial behavior,* and *marital disruption.* Code type and family history were not so strongly related as delinquency and family history. However, signs of intra-family psychological distress and dissocial behavior by other family members go with boys who had 489 codes. This relationship was most apparent in the period between 16 and 21 years, the post-testing period.

Information Gathered from Parents

In this section, the findings obtained from the History Records (see Footnote 2) completed by the parents of *S*s are reported.

Sampling Success

One hundred and sixty-three records, 58% of the total sample, were completed. Eleven History Records were completed by family members other than an *S*'s mother. The number of records received for each sample appears in Table 14. One hundred and forty-three *S*s had both the interview and the History Record, while 20 *S*s had only a History Record, and 54 cases had only an interview. Seventy-nine cases had neither an interview nor a History Record.

The samples were divided into the *S*s represented and not represented in the information gained from their parents. These groups were then compared using the findings from the Hathaway-Monachesi ratings, the Social Agency Survey, and the delinquency rating. It was found that (*a*) the delinquent samples were not as well represented as the nondelinquent samples, a difference which seems to be due largely to the more frequent deaths of the delinquent boys' mothers; the degree of severity in the represented and nonrepresented samples was not different; (*b*) the unrepresented cases had been rated in 1952 as less well educated and less disposed to constructive use of time and participation in community activities; (*c*) in the Social Agency records, the unrepresented cases were recorded as having in Sample 2 more *Poverty and Economic Need,* in Sample 3 more *Psychiatry for the Family,* while the represented cases had more *Dissocial Behavior in the Family.* In Sample 4 there were four categories reflecting uneven sampling: *Family Disruption Due to Disease, Poverty and Economic Need, Dissocial*

Behavior in the Family, and *Inadequate Parent-Child Relationships.* All of these were more frequent in the sample without history records from their parents.

The differences noted above showed more socially desirable characteristics to predominate in the samples with History Records, which suggests that comparisons concerning these variables may be considered minimal estimates of the true situation in the population. Better sampling would almost surely have increased the number and significance of relationships found.

The 138 items in the History Records were analyzed for differences between codes (Samples 1 plus 2 vs. Samples 3 plus 4), within codes (Sample 1 vs. Sample 2 and Sample 3 vs. Sample 4), and for delinquency (Samples 1 plus 3 vs. Samples 2 plus 4). The rating of severity that was used is the rating made in 1956. The mild delinquents were those with ratings of 4 points or less on this total scale; the most severe rated 5 to 9 points.

The results showed 35 items with differences at the .05 level or less. These items were arranged under three topics (Tables 15, 16, and 17) covering family background and family characteristics, education-occupation and social history, and personality characteristics.

Certain trends are immediately obvious in the 24 items differentiating the delinquents from the nondelinquents (Sample 1 plus 3 vs. Sample 2 plus 4). Eighteen items occurred again in the comparison of Samples 1 and 2, while only three were reflected in reliable differences between Samples 3 and 4. This last is probably an aspect of the highly selected cases in Sample 4 that were examined. There were only seven items differentiating between code groups.

Family Background and Family Relationships. In the area of family background and family relationships (Table 15), differences were mainly between delinquents and nondelinquents within 025 code cases. These items suggested that the homes from which the boys in Sample 2 came were economically poorer (unsteady income, absence of home ownership). The Sample 2 boys' parents made less acceptable adjustments in their relationship to the law. In the boys' relationships with their parents there was more conflict (in the form of dis-

TABLE 14

CASES WITH HISTORY RECORD AND INTERVIEW DATA IN THE FOUR SAMPLES

Information	Sample								
	1	2	Mild	Severe	3	4	Mild	Severe	Total
Interview and history record	44	33	19	14	36	30	15	15	143
Interview only	9	15	5	10	8	12	4	8	44
Not interviewed			16	7			14	17	
Dead	2	2	1	1	1	2	1	1	7
Moved or lost									
before 1952	3	5			6	3			17
1952-56	4	1			3	5			13
In service and otherwise unavailable for interview	3	2	1	1	9	2	0	2	16
Refused	3	10	7	3	6	12	5	7	31
Not located, 1956	3	3	7	2	2	7	8	7	15
History record only	5	3	2	1	8	4	0	4	20
No data (history record and interview)	13	20	14	6	19	27	14	13	79
Sample *N*	71	71	40	31	71	73	33	40	286

agreements about coming in late and fear of the mother).

The code differences in family relationships suggested that the 025 boys came from homes that were more permissive or more sophisticated, or perhaps more *laissez-faire.* (Parents seldom punish and more often allow the children to taste alcohol.)

Educational and Social Adjustment. In the area of educational and social adjustment (Table 16) the delinquents and the 489 code type boys showed less educational success. (They quit school earlier as seen in the Hathaway-Monachesi data.) The delinquents (alone) made an earlier adjustment to the role of wage earner and found a position at a lower social level than the nondelinquents. In the service they tended not to rise in rank.

In the more purely social areas, delinquents had fewer formal social contacts and in those that they had, they less often attained leadership status. Their friends were more often of a lower social status than their own families. They tended to meet their friends in public places and not at home. Finally, they expressed less interest in religion.

Personal Characteristics. In their personal characteristics (Table 17), the delinquents were more excitable, more often day dreamers, and more often showed poor judgment. They were seldom sick.

The 489 delinquents also were less often sick, drank more, and fought more when drinking.

Other items are interesting although they are not so easily grouped. The fact that the severely delinquent boys' parents were more often out of the home early, if at all, suggests the importance of the presence of the parents.

TABLE 15

PERCENTAGES FOR ITEMS FROM THE HISTORY RECORD CONCERNING FAMILY BACKGROUND AND FAMILY RELATIONSHIPS THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
1	Homes owned by parents	47	65			69	39				
13	Family income not steady	51	33			35	61				
9	Parents left their homes when *S* was young if left at all			76	50						
20	Family members in trouble with law	22	06			4	28				
128	Father spent time in prison	11	04			0	11				
75	Disagreements with parents	83	67			61	92				
87	Trouble with late hours	40	16					18	44		
26	Seldom punished									09	26
30	Sometimes afraid of mother	34	22			22	33				
82	Often jealous as a child	32	18			16	39				
72	Allowed to taste alcohol as a child									38	54

Findings from the interview

In this section, the results of the clinical interview (see Footnote 2) conducted with the *S*s are reported.

Sample Representation

The number of cases interviewed and the six different categories of cases not interviewed were presented in Table 14. Of the 286 individuals in the total sample, 187 (two-thirds) were interviewed. There was no reliable interaction between the number interviewed and delinquency or code group. There was no reliable interaction between the four samples and the over-all frequency of cases interviewed and cases not interviewed.

The cases not interviewed were compared with those who were interviewed on the Hathaway-Monachesi ratings, the social agency findings, and the delinquency ratings. In terms of total sampling, the interviewed cases in contrast to the total group were boys who cooperated or were available and who were a little older on the average, and slightly better educated; they had better social adjustment in that they made more constructive use of leisure time, and displayed greater interest and achievement in community activities. Among the delinquents interviewed, no difference in the average severity of delinquency from those in the samples which they represent was found, though their families have had fewer troubles in the areas of disease and poverty. It seemed appropriate, therefore, to make

TABLE 16

PERCENTAGES FOR ITEMS FROM THE HISTORY RECORD CONCERNING EDUCATIONAL, OCCUPATIONAL, AND SOCIAL HISTORY THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
31	Quit school at early age	66	43			29	61			57	37
35	Don't go on to school (college, special training, etc.)	43	27			18	39			41	27
76	Often in trouble in school	62	38			29	59				
38	Earned money before 15 years of age	67	58			51	78				
39	Obtain a lower job level (skilled, unskilled)	67	38			26	69				
119	Rank in service of lowest order, pvt.					14	36				
63	Friends often from lower or different social status than *S*	42	22			22	47				
64	Have few club or lodge contacts			62	53						
65	Seldom are club officers	92	81			27	67				
62	Enjoys self most in public places					14	33				
126	Attends church once or less than once a month	62	44			34	66				
127	Reads few books			30	14						

the same assertion concerning the sampling of interviewed cases as was made about the results of the History Records in the preceding section. That is, any differences found among the samples on the basis of the interview data which reflect more socially undesirable characteristics in one group or another represented a minimal estimate of the probable relationship in the parent populations.

Differences Found with the Clinical Interview

The distributions of scores from the interview items were compared for the three variables: delinquency, severity of delinquency, and code type. The distinction between mild and severe delinquents was again made on the basis of the 9-point rating of delinquency though 1956. Mild delinquents had four or less points, while severe delinquents had more than four points.

The results from the interview are presented in Tables 18, 19, 20, 21, and 22. Items with significant differences at the .05 level or less were grouped a priori into five topics.

Family Background and Family Relationships. Differences pertaining to the first area, family background and family relationships, are presented in Table 18. This table shows the percentage of *S*s in each group responding to items which significantly discriminate between groups. There are 14 items differentiating between the delinquent and nondelinquent groups. Three of these items also differentiated between the mild and severe delinquents. Only three items reflected code group differences (Sample 1 and 2 vs. Sample 3 and 4).

The delinquents' fathers were not so well educated, had more trouble with the law, and drank more frequently than the nonde-

TABLE 17

PERCENTAGES FOR ITEMS FROM THE HISTORY RECORD CONCERNING PERSONALITY CHARACTERISTICS THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
99	Excitable and enthusiastic	34	22								
91	Sometimes a daydreamer	87	74			69	89				
74	Mistakes in judgment	50	22			18	50				
77	Sometimes spends money foolishly	66	46			35	68				
85	Had sweet tooth as a child	46	28								
88	Doesn't pay attention to what he eats									9	21
92	Drinks alcoholic beverages							86	63		
45	Had more fights when drinking	21	4	23	9	0	19	7	24		
112	Not sick as a child							81	59		
113	Not sick as an adolescent	75	92					95	70		
107	Often goes to doctor									22	39
124	Tells funny stories									22	39

TABLE 18

PERCENTAGES FOR ITEMS FROM THE INTERVIEW CONCERNING FAMILY BACKGROUND AND FAMILY RELATIONSHIPS THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
39	Father had (no more than) 9 years of schooling	59	41								
42	Father drank daily	43	24	35	10						
43	Father had minor offense to serious trouble with law	33	14			6	39				
51	Mother drank irregularly									60	44
46	Mother had unskilled jobs	28	11					15	35		
52	Mother sometimes gone from home	6	16								
55	Did not associate with siblings	69	48					42	68		
57	Sibling relationships smooth and easy	39	21							39	21
56	Siblings got into minor trouble	29	15								
59	Raised in inadequate home	25	11								
58	Raised in multiple-dwelling unit	30	15								
62	No conflicts over adolescent behavior because of permissiveness	26	16								
62	Parental concern over usual or unusual behavior	51	33								
100	Center of activities away from home	53	31	64	41	43	69			48	24
64	No longer close to family	29	13	27	15	9	21	18	35		

Note.—All differences are as rare or rarer than 1 in 20.

linquents' fathers. The delinquents' mothers held lower-level jobs and were out of the home more. The home itself was more often considered by the *S*s to be physically inadequate, and was more often a multiple-dwelling unit.

In his relationships with his parents, the delinquent elicited concern over both usual and unusual behavior, but he had less conflict with his parents because they were permissive. He associated less with his siblings, but said that his association with them were smooth and easy more often than did the nondelinquent. At the time of interviewing, the delinquents were less involved with their families, which follows from the trend for the delinquents' homes to be less a center of activities than for the nondelinquents.

Three Items Related to Code. The 489 cases report that their mothers drank irregularly, while the 025 cases indicated that their mothers drank less if at all (Sample 3 plus 4 vs. Sample 1 plus 2). The sibling relationships of the 489 group were said to be smooth and easy, and the center of activities was away from home for them.

Education and Occupation. The items describing educational and occupational differences between samples appear in Table 19. Again differences between the delinquent and nondelinquent groups showed differences within both of the code groups.

TABLE 19

PERCENTAGES FOR ITEMS FROM THE INTERVIEW CONCERNING EDUCATIONAL AND OCCUPATIONAL HISTORY THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
7	No more than high school education	69	39			31	67	48	71		
9	Either disliked school completely or liked just a few aspects of school	33	16								
11	C student	80	58			56	88				
13	No more than 2 extra-curricular activities in high school	83	67								
16	No recognition for performance	67	47								
99	Leisure time wasted in high school	44	28			38	60	39	71		
12	Often in trouble in school and expelled at least once	29	6	27	9			12	43		
22	Full time job before age 18	62	48			40	67				
23	Had 2–11 full-time jobs	68	50								
27	Changes jobs when opportunity arises			80	57						
24	Skilled, semi-skilled, unskilled worker	78	58			40	66				
28	Seeks semi- or skilled jobs	21	10								

Note.—All differences are as rare or rarer than 1 in 20.

On the whole, the delinquents were less successful at school. They received lower grades and participated in fewer extracurricular activities, stayed in school a shorter time, often not going beyond high school, disliked school completely (even at the time of the interview), or liked just a few aspects of school. The delinquents received less recognition for their school performance, and were in trouble and were expelled more often than were the nondelinquents.

Occupational pursuits came earlier to the delinquents who frequently held a full-time job before the age of 18. The delinquents sought and held lower-level jobs and held many more jobs than the nondelinquents.

The delinquent 025 sample appeared to be more clearly below its nondelinquent code pair (Sample 1) in academic grades, job level, and age of full-time job than was the delinquent group with 489 code (Sample 4) below its nondelinquent mate (Sample 3).

Social Patterns. The items relating to social patterns are presented in Table 20. The majority of differences were between Samples 1 and 2. Thirteen items reflect differences between those samples. The delinquent groups had left home, and were less involved with their family. They tended more often to have friends who were married, who were from a lower social level, and more often had been delinquent themselves than had the nondelinquents' friends. Contact with friends was more often in places of public entertainment among the delinquents.

The delinquent himself was considered extroverted in action, adventurous, valued entertainment highly, and led in social activities more often than the nondelinquent.

TABLE 20

PERCENTAGES FOR THE ITEMS FROM THE INTERVIEW CONCERNING PRESENT SOCIAL PATTERNS THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
4	Has left parental home	61	42								
100	Interests lie away from family	66	46	48	24	43	69				
106	Friends are of different or lower social levels	57	37								
107	Friends contacted in places of public entertainment	33	16	22	12	10	39				
108	Friends have delinquency records	72	26			25	69	27	74		
114	Friends are married	49	27			24	62				
112	Friends are skilled or unskilled workers	58	24			23	66				
109	Leads in social activities	22	9								
114	Goes to church no more than once a month	70	50					56	69		
	Church is unimportant or of moderate interest	67	43	94	79	34	61			83	71
117	Value entertainment	29	14			2	25				
120	Adventurous	49	20			9	42	34	57		
121	Extroverted in action	43	26			17	37				

Note.—All differences are as rare or rarer than 1 in 20.

The delinquent also expressed less interest in religion and went to church less often.

These differences were noted earlier and appeared to lie between the delinquents and nondelinquents in the 025 code group (i.e., Samples 1 and 2).

Marital and Sexual Patterns. The items describing marital and sexual patterns appear in Table 21. Again, these differences were largely between Samples 1 and 2. The delinquent group was more precocious, promiscuous, and active with respect to pre- and postmarital sexual relations. The two 489 samples did not respond differently to the items in this area.

Trouble. The items differentiating between the delinquent and nondelinquent samples that describe trouble are presented in Table 22. The delinquents in both code groups were different from their nondelinquent code pairs on these items: drinking, trouble with drinking, and trouble in service. A more extensive presentation of the actual offenses committed was given earlier in Table 5.

Special Sets. Four sets of items were included to test specific hypotheses. The first was ten items taken from the Gluecks' work. In Table 23 the delinquents and nondelinquents are compared within the two code groups. Six statements differentiated between the total delinquent and nondelinquent groups (Samples 2 plus 4 vs. Samples 1 plus 3). Two items that show significant differences referred to the home (the father's discipline and family cohesion). Four items described personality: adventurous, extroverted in action, stubborn, and emotionally unstable. Individual code group differences are shown in Table 23. The severe and mild delinquent groups differed in mother's supervision and father's affection. A detailed study of the predictive utility of these items will be published separately. The three hypothetical a priori clusters, "*Pd* signs," "Opportunities to Act Out," and "Changes in Life Adjustment after the 9th Grade," all contained items of such low frequency that they did not permit statistical investigation.

Findings from the Q Sort

The Block *Q* deck was sorted at the end of each interview. The 115 items were dis-

TABLE 21

PERCENTAGES FOR ITEMS FROM THE INTERVIEW CONCERNING SEXUAL PATTERNS AND MARITAL PATTERNS THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
79	Sexual intercourse before age 18	62	30			34	77				
80	Sexual relations with a few or many women who were not his wife	80	52	90	68	40	73			78	56
81	Pre-marital sexual relations	74	46			37	71	57	79		
81	Sees prostitutes	22	8			8	23				
70	Never engaged	17	41			43	17				
77	Married	58	37			34	56				
83	Sexual relations no more than once a week	38	63					66	38		

Note.—All differences are as rare or rarer than 1 in 20.

tributed into a forced 9-point distribution. In analyzing these results, the items were interpreted individually. For each sample item, the mean and *SD* were calculated. A detailed study of stereotypes using this *Q* deck will be published separately.

The results showed differences between the four means on each of the 115 items. Forty-one items showed sample differences at the .05 level or less. The 41 items included 65 differences between samples. A shortened form of the items and differences is reported in Table 24. It will be recalled that these items are actually ratings of each case as to whether the statement describes (is like) or does not describe (is unlike) the individual. The sample means were compared on each rating. This has resulted in clusters of items each responding in the same way in the comparisons. In Table 24,

TABLE 22

PERCENTAGES FOR ITEMS FROM THE INTERVIEW CONCERNING DELINQUENCY AND TROUBLE THAT RELATE TO DELINQUENCY, SEVERITY, AND CODE TYPES

Item No.	Content of Sample Differences	Reliably Different Percentages									
		D vs. ND Groups		Severe vs. Mild		Sample 1 vs. 2		Sample 3 vs. 4		Code Types 489 vs. 025	
84	Calls self delinquent	70	20	64	28	00	14	2	19		
89	Drinks daily to excess at times	47	28								
94	Trouble while drinking	28	6			2	19	8	41		
33	Military disciplinary trouble	12	2								

Note.—All differences are as rare or rarer than 1 in 20.

TABLE 23

DELINQUENTS AND NONDELINQUENTS COMPARED ON TEN ITEMS TAKEN FROM THE GLUECKS'[a] STUDY OF DELINQUENCY

(All differences are in the expected direction)

Item No.	Content of Item	Significant Differences			
		D vs. ND Samples 2 and 4 vs. Samples 1 and 3	Sample 1 vs. 2	Sample 2 vs. 4	Mild vs. Severe
74	Father's discipline	.01	.01		
75	Father's affection		.02		.02
76	Mother's supervision				.01
77	Mother's affection				
78	Family cohesion	.05			
120	Adventurous	.01	.01	.05	
121	Extroverted in action	.02	.02		
122	Suggestible				
123	Stubborn	.05			
124	Emotionally unstable	.02			

[a] Glueck and Glueck, 1950.

TABLE 24

Q-Sort Items That Differentiate between the Samples

Samples Less Like the Item	Sample More like the Item			
	1	2	3	4
1	90. Philosophical 79. Has obsessive thought 7. Conservative 9. Disturbed by uncertainty *Characteristic*[a]	56. Responds to humor 50. Unpredictable 93. Masculine 27. Adequate sexually	28. Accepted by people 57. An expressive person 88. Charming 27. Adequate sexually	53. Acts out 62. Nonconforming 85. Communicates by doing 50. Unpredictable 93. Masculine 23. Extrapunitive 80. Seeks opposite sex 96. Values autonomy 94. Expresses hostilities 73. Perceives things in sexual terms 58. Enjoys sensory contacts 67. Self-indulgent
2	25. Represses needs and impulses 38. Is sarcastic 41. Moralistic and strict 10. Tensions converted into somatic symptoms		91. Competitive 57. An expressive person 88. Charming	89. Envious 67. Self-indulgent
3	106. Has anxiety 87. Complicates simple situations 10. Tension converted into somatic symptoms 59. Concerned with body	9. Disturbed by uncertainty		53. Acts out 62. Nonconforming 85. Communicates by doing 37. Deceitful 1. Critical 69. Sensitive to demands 30. Gives up when frustrated
4	51. Values intellectual matters 24. Likes being rational 25. Represses needs and impulses 38. Is sarcastic 41. Moralistic and strict 14. Submissive 59. Concerned with body	40. A worrier	78. Is orderly 8. Has intellectual ability 14. Submissive	

[a] Items which differentiate between Group 1 and all other groups.

an item characterizes the sample under which it is reported and is atypical of the sample listed to the left. Thus each item is entered into the cells which represent the difference between the means for the item. The cells on the major diagonal of the matrix were reserved for items differentiating each sample from all other samples. Only Sample 1 has items on this diagonal.

Most differences occur in the first and fourth colums in Table 24. There were four items characteristic of Sample 1 alone. The differences between Sample 1 and the other samples were numerous, as were the differences between Sample 4 and the nondelinquent samples (Samples 1 and 3). Sample 1 was said to be more philosophical, obsessive, conservative, and disturbed by uncertainty than all of the other samples. In comparison to Sample 2, Sample 1 was more likely to repress impulses and to convert tensions into somatic symptoms, was more materialistic, and was more sarcastic. Sample 2 was described as more unpredictable, masculine, sexually adequate, and responsive to humor than Sample 1. Sample 1 was said to have more anxiety than Sample 3, to be more retentive, to complicate simple situations, and to convert tensions into bodily symptoms. Sample 3 was said to be more accepted by people, more expressive, more charming, and more adequate sexually than Sample 1. Sample 1, in comparison to Sample 4, was said to value intellectual matters more, to repress needs and impulses, to be more sarcastic, moralistic, submissive, and concerned with the body. Eleven statements contrast Samples 2 and 4. Sample 4 exceeds Sample 2 in mean rating on nonconforming, communicates by doing, unpredictable, masculine, extrapunitive, acts out, values autonomy, expresses hostility, self-indulgent, enjoys sensory contacts, perceives things sexually. Sample 4 was said to be more nonconforming than Sample 3, to act out more, to be deceitful, critical, sensitive to demands, and to give up when frustrated. Sample 3 was said to be more orderly, intellectual, and submissive than Sample 4. In general, one might interpret these results as showing Sample 1 (nondelinquent 025's) as appearing to be the most neurotic of the four samples, Sample 4 (delinquent 489's) as the most delinquent, Sample 3 (nondelinquent 489 s) as the best adjusted, and Sample 2 (delinquent 025's) as the most inconsistent and heterogeneous, an atypical group, difficult to categorize.

Item Clusters from the History Records and the Clinical Interview

In this section, some items from the History Records and from the interview are examined as clusters. In such fields as psychiatry, clinical psychology, industrial psychology, and counseling, the individual case material frequently is more easily apprehended in groupings, clusters, trends, or outlines of events or behavior (Briggs, 1955).

In the earlier work by Briggs (1955; in press) objective data were grouped into well-recognized a priori clusters, after which statistical attributes of reliability and validity were examined. In an investigation such as this, the selection of clusters for examination is analogous to the initial phase of conducting an interview.

TREATMENT OF HISTORY RECORD CLUSTERS

Three clusters isolated by Briggs in his original study had relevance for the population studied here. Comparisons for the four samples are reported for clusters of items describing *Achievement, Psychopathy,* and *Conflict with Parents.* The items in the original set taken from the M-B History Records were refined by the item analysis technique suggested by Loevinger and DuBois (1953). The scoring of each cluster involved first the distribution of scores from each item for all of the four sample distributions. Weights for each item in the frequency distribution were devised so that the distribution had an *SD* of 10 and a mean of 50. These 20 scores were then used to weight each alternative. Each individual *S*'s record was scored by summing the scores for the items in each cluster. Items unanswered were arbitrarily assigned a score of 50, the mean of the entire sample.

The distribution parameters examined were the means and variances. The data for each cluster were examined through the analysis of variance technique, when the data met the required assumptions of variance homogeneity. The means and variances for the samples are also presented for inspection. Cluster scores not conforming to the assumption of the variance analysis technique were examined with *t* ratios, and were sufficiently in excess of appropriate criteria to make unnecessary a more exact test.

Treatment of Interview Clusters

The interview was both broader and more specific in its coverage of some family characteristics than was the History Record. Two aspects of the boys' families were considered of special interest, and clusters of items were developed for these. The clusters were called *Sufficiency of the Family* and *Family Occupational-Educational Achievement.* Scoring for each of these sets was dichotomous. No attempt was made to weight the entire distribution of points, as was done with the clusters taken from the History Record. These two clusters were measured in the interview data only, and the sample means and variances were compared. A single index of reliability, the Hoyt (1941) analysis of variance techniques was calculated for each of these scales. The sum of squares was partitioned into the quantity attributable to the items, *S*s, interaction between items and individuals, and to the residual. The *F* tests permitted comparison of the residual (error) variance and the *S* variances, as a check on the meaningfulness of each cluster. Similarly, a reliability coefficient was obtained. This latter value is of minor interest since the item clusters were not compared with one another. The major interest was in comparisons among code and delinquency groups.

Achievement. The items for this cluster are listed in Table 25. The items listed here describe school and job success almost exclusively. Two of the items already have been shown to discriminate between delinquent and nondelinquent groups; these findings are summarized in Column three of Table 25.

The means and variances of the four samples on the nine Achievement items are presented in Table 26. The table also shows the *t* and *F* ratios resulting from a comparison of these values. This table shows the variance of Sample 3 to be signficantly smaller than the variance in the other samples. This is evident from the *F* ratios reported. On the *L* test (Bartlett's test [Johnson, 1949, pp. 83-84]), one would conclude that with respect to variance, the samples were drawn from different populations ($P < .05$).

Higher values indicated greater levels of achievement. Comparison of the means in Table 26 shows that Sample 1 achieved more than all other samples. Samples 2 and 4 were indistinguishable, while Sample 3 was little different from these delinquent samples when compared to Sample 1. Thus, in terms of school and job success, Sample 1 achieved highly, and the other three groups did not.

Psychopathy. The 13 items defining this concept are presented in Table 27, with the six items indicated for which the sample differences were found.

The analysis of variance results for the *Psychopathy* cluster are presented in Table

TABLE 25

ITEMS FOR THE CONCEPT OF ACHIEVEMENT FROM THE HISTORY RECORD

(Group or sample differences associated with each item are reported)

Item	No.	Group Differences
Father's occupation	(21)	
Amount of *S*'s education	(31)	ND > D1 > 2
Progress in school	(32)	
Difficulty with 3 r's	(33)	
Grades	(34)	
College or further training	(35)	ND > D1 > 2; 1 and 2 > 3 and 4
Interest in school	(37)	
S's best vocation is ______	(39)	
School and job success	(42)	

Note.—Numbers in parentheses refer to History Record number.

TABLE 26

ACHIEVEMENT, SAMPLE MEANS, AND VARIANCES WITH SIGNIFICANT DIFFERENCES BETWEEN SAMPLES ON t TESTS AND F RATIOS[a]

Group	Group 1	2	3	4
1				
2	5.34 (2.06)			
3	4.78 (2.61)			(1.83)
4	4.77			
Mean	541	483	492	483
Variance	3605	1754	1381	2527
N	49	36	44	34

[a] F ratios are numbers within parentheses; t ratios are not in parentheses.

28. Sample means and variances are given in Table 29.

No reliable variance differences were suggested in the Bartlett's test. The results of the variance analysis demonstrated that *Psychopathy* as measured in this cluster was associated with delinquency. Also, there appeared to be a definite association between the amount of *Psychopathy*, the code type, and delinquency. Thus, the Sample 2 mean was much higher and the Sample 1 mean much lower than one would expect. That is, within the 025 code groups, the highest *Psychopathy* cluster scores of any sample were obtained by the delinquents with an 025 personality test pattern, while the lower *Psychopathy* cluster scores were obtained by the nondelinquents having the same personality test characteristics.

Conflict with Parents. The ten items describing the cluster of *Conflict with Parents* are listed in Table 30. These items seem to reflect a tendency for the parents' treatment to alienate the *S*s. Three of these items, when taken individually, were highly associated with delinquency on the part of the *S*s. The results of the analysis of variance model are presented in Table 31. The group means and variances are given in Table 32.

The Bartlett test of sample variance homogeneity suggested that it was permissible to obtain a pooled variance estimate. The association between conflict which leads to alienation of parents and the boys' later delinquency was undeniably shown. Again, the interaction between code and delinquency suggested differences among the 025 code-type samples. The delinquent 025 boys appeared to come from much more rejecting homes than was true of the homes of

TABLE 27

ITEMS FOR THE CONCEPT OF PSYCHOPATHY FROM THE HISTORY RECORD
(Group or sample differences associated with each item are reported)

Item	No.	Group Difference
Family trouble with law	(128-20)	D > ND 2 > 1
Interest in school	(37)	
Promiscuity	(44)	
Type of punishment in childhood	(53)	
S's judgment	(74)	1 better than 2; ND better than D
Behavior in school	(76)	1 better than 2
Judgment in money matters	(77)	1 better than 2; ND better than D
Temper of *S*	(86)	
Family attitude toward late hours	(87)	ND better than D; 3 better than 4
Amount of drinking by *S*	(92)	3 < 4
Job mobility	(97)	
S's closeness to family now	(99)	

Note.—Numbers in parentheses refer to History Record Numbers.

the nondelinquent boys having similar personality test profiles (Sample 2 vs. Sample 1).

The results of the group comparisons showed that the concept of *Conflict with Parents* closely paralleled the findings with the cluster of *Psychopathy*. Only two items were the same on the two scales, however, which suggested that while the phenomena were correlated, the correlation was not due to item overlap.

TABLE 28

ANALYSIS OF VARIANCE OF SCORES FROM THE CLUSTER PSYCHOPATHY

Source of Variation	*df*	Variance	*F*
Between samples	3	28443.33	
Code	1	4930.75	
Delinquency	1	58843.0	17.01**
Code × Delinquency	1	21556.75	6.23*
Within samples	159	3458.33	

* $P < .05$
** $P < .01$

TABLE 29

SAMPLE MEANS AND VARIANCES FOR SCORES FROM THE CLUSTER PSYCHOPATHY

Sample	*N*	Mean	Variance
1	49	612	2804
2	36	673	4413
3	44	646	2658
4	34	661	3976

Interview Clusters

Family Sufficiency. This cluster was designed to include all items that reflected upon the adequacy of a family as viewed from conventional standards. The items are presented in Table 33 with the scoring directions.

An analysis of the items using the Hoyt reliability technique suggests that the *S* scores greatly exceeded the residual variance, which means that the items were operating in a correlated pattern. The reliability of the item set was .84. The pattern of intra-scale variation indicated that analysis of the items as a cluster was justified. The analysis of variance of these scores appears in Table 34. The means and variances for the four samples are given in Table 35.

The variance analysis was shown to be legitimate since the variances were not significantly different on Bartlett's test. The pattern demonstrated was clear; the delinquents came from much less sufficient homes than the nondelinquents.

Family Occupational-Educational Achievement. This cluster of seven items

TABLE 30

ITEMS FOR THE CONCEPT CONFLICT WITH PARENTS FROM THE HISTORY RECORD
(Group or sample differences reported for individual items)

Item	No.	Group Differences
Frequency of punishment	(26)	3 and 4 > 1 and 2
Frequency of parent's quarrels	(27)	
Afraid of father	(29)	
Frequency of bringing girls home	(48)	
Number of times family denied dating of particular girl	(51)	
Type of punishment in childhood	(53)	
Frequency of bringing fellows home	(61)	
Amount of conflict with parents	(75)	ND < D; 1 < 2
Amount of conflict about smoking and drinking during adolescence	(80)	
Family attitude toward late hours	(87)	ND better than D; 3 better than 4

Note.—Numbers in parentheses refer to History Record items.

TABLE 31

Analysis of Variance for Scores on the Cluster Conflict with Parents

Source of Variation	*df*	Variance	*F*
Between samples	3	12350.58	
Code	1	5868.00	
Delinquency	1	13203.00	7.16**
Code × Delinquency	1	17890.75	9.71**
Within samples	159	1842.73	

** $P < .01$

from the interview described the family level or education, occupation, and home adequacy. They are given in Table 36.

The Hoyt variance analysis suggested that further analysis of the set as a whole was valid. The reliability of this set was .64. An analysis of variance of *S*'s scores in the samples is presented in Table 37. The single effect of this analysis was the demonstration of the association of low family occupational-educational achievement and delinquency. The sample means and variances are given in Table 38.

Cluster Interrelation

The final examination of the five clusters discussed above involved their interrelationships within the samples. The intercluster correlation matrix for each sample appears in Table 39. For the two clusters drawn from the interview, *Family Sufficiency* and *Occupational-Educational Level,* the number of cases was the same as the number interviewed. Similarly, interrelationships between History Record clusters were based upon the number of cases with a History Record. The correlation coefficients between the interview and the History Record relied upon the overlapping cases ($N = 143$), noted in Table 14.

TABLE 32

Sample Means and Variances for Scores from the Cluster Conflict with Parents

Sample	*N*	Mean	Variance
1	49	477	1798
2	36	516	2149
3	44	510	1979
4	34	507	1448

The interpretation of these interrelationships between home factors and *S* behavior pattern is interesting from an explanatory viewpoint. It is already known that four of the clusters were sensitive to delinquency. Also shown were the essentially similar variances within the samples. *Achievement* was the only exception.

The most striking single factor apparent in Table 39 was the contrasting size of the coefficients in the delinquent and nondelinquent samples. There were nine coefficients which exceeded the chance level ($P < .05$) in the nondelinquent samples, while in the delinquent group there was but one. From the pattern of relationships between the samples, a number of trends appeared which suggested that other coefficients in larger samples would probably be significant. Statistical artifacts (distribution shape) often control the size of the correlation coefficient, but the Bartlett's tests conducted above leave little encouragement for the argument that such was the case in these distributions. From the psychological standpoint these results suggested that there was more intra-home variability in the delinquent samples.

The strongest relationship in these matrices is between *Psychopathy* and *Conflict with Parents.* In the nondelinquent samples there is a relationship between *Conflict with Parents* and *Family Sufficiency. Psychopathy* and *Family Sufficiency* also relate noticeably, forming a triad that appears to relate family behavior with children's behavior. A second set of clusters that carry a heavy relationship is the *Occupational-Educational Level* of the home and the *Achievement* level of the son. To this pair may be added the *Family Sufficiency* and *Occupational-Educational* relationship. These concepts distinguish home patterns of delinquents or nondelinquents. They are not highly correlated.

The means for the five clusters of items described above are presented in Fig. 3.

Each sample mean was calculated from a distribution of *z* scores which for the four samples had a mean of 50 and a standard deviation of 10. (This represents a translation of the means of the values in Tables 26, 29, 32, 35, and 38). This permits both intersample and intrasample comparisons. The two scales, *Psychopathy* and *Conflict with Parents,* are inverted in Fig. 3 for clarity of exposition.

The three kinds of characteristics are subsumed in the group and scale variables in Fig. 3. The initial reference points are the family characteristics. There were two of these, *Family Sufficiency,* a conformity and morality composite, and *Family Occupational-Educational Level.* Two mediating concepts are represented. These are the code classification of the sample and the History Record cluster and *Psychopathy.* Both concepts are dispositional personality variables coming before the consequent behavior. Finally, there are three classes of behavior that we may seek to understand. These are (*a*) delinquency, (*b*) the sample classification variable, and (c) *Achievement* and *Conflict with Parents.*

From Fig. 3, five generalizations may be made relating to these concepts. These rely upon the relative position of the means. (*a*) Delinquency was a function of the two family characteristics. The samples with better scores on the family characteristics were not delinquent; those with lower means on these characteristics were delinquent. This result merely echoes numerous points made

TABLE 33

ITEMS FOR THE CONCEPT SUFFICIENCY OF THE FAMILY DEFINED BY ITEMS FROM THE INTERVIEW
(Significant group or sample differences reported for individual items)

Sufficiency of the Family Item	Points of Dichotomy		Group Differences
	1	0	
Family status (35)	normal	other	
Father's earned income (41)	steady	other	
Father's drinking (42)	irregular	daily, now and then	ND > D; Mild > Severe
Trouble with the law (43)	minor	serious	ND > D
Father's absence from home (44)			
birth to 1 year	(1)	a	
2-9, 16-20 years	(2)	a + b	
10-15 years	(3)	a + b + c	
multiple	(4)	a + b + c + d	
Mother's drinking (51)	never	irregular	1 and 2 > 3 and 4
Mother's absence from home (52)			
16-21 years	(1)	a	
10-15 years	(2)	a + b	
0-9 years	(3)	a + b + c	
Relation with siblings (55)	close	not close	ND > D; 3 > 4
Trouble with law by siblings (56)	minor	serious	ND > D
Type of dwelling (58)	owned	rented	ND > D; 3 > 4; 1 > 2
Parental quarrels (60)	occasional	frequent	Mild better than severe, 1 better than 2; ND better than D
Religious atmosphere (61)	religion	no religion	ND better than D; mild better than severe
Family cohesion (63)	normal	none	
Mental illness in family (98)	none or distant relatives	in immediate family	

Note.—Numbers in parentheses refer to the interview item number.

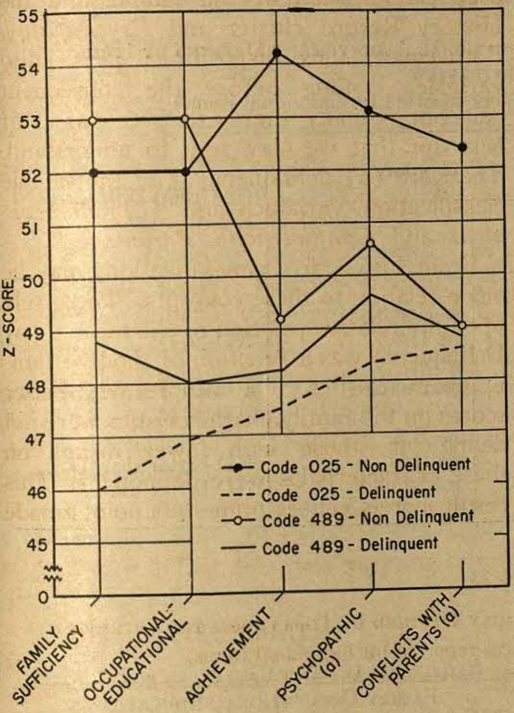


FIG. 3. Sample means on five clusters presented in z-score form. (Scale inverted for clarity of exposition.)

earlier. (*b*) Delinquency was a joint function of *Psychopathy* and *Conflict with Parents*. The samples in which the *Psychopathy* mean exceeded the mean on *Conflict with Parents* were delinquent while those in which the reverse was true were not delinquent. This result suggests that the groups who acted out their family conflicts in all areas got into trouble, while those who could discriminate between parents and society at large did not. (*c*) Delinquency was a function of *Psychopathy* alone; more psychopathic samples were delinquent and the less psychopathic samples were not. (*d*) *Achievement* was an inverse function of *Psychopathy*. The order of sample means on one of these concepts is the reverse of the other. (*e*) *Achievement* appeared to be a joint function of code and family characteristics, whereby the achievement level in the 025 samples was above the average for the two means of family characteristics; while in the 489 samples, the *Achievement* was below the average of the family characteristics.

TABLE 34

ANALYSIS OF VARIANCE FOR THE CLUSTER SUFFICIENCY OF THE FAMILY

Source of Variation	*df*	Variance	*F*
Between sample	3	47.289	
Code	1	14.055	
Delinquency	1	125.309	12.49**
Code × Delinquency	1	2.505	
Within samples	183	10.09	
Total	186	1988.499	

** $P < .01$

Certain interesting trends are apparent from the data of Fig. 3. Probably the most important is a re-evaluation of Sample 3. The level of achievement demonstrated by this sample was so low that it suggests that these boys were making a far from satisfactory adjustment—much less satisfactory than their nondelinquent status would indicate. In the context of concern over human resources, there is probably more waste involved in this combination of background and personality than in the other samples. Thus, in comparison, the miserable origins of Sample 2 with the 025 code still did not lead to a level of performance below their own family level as is seen in the delinquent 489 samples.

A second point of interest is the relationship between *Psychopathy* and *Family Conflict*. The nondelinquent combination of these two variables is the adjustment that one would predict as most satisfactory either from a viewpoint of dynamic psycho-

TABLE 35

SAMPLE MEANS AND VARIANCES FOR SCORES FROM THE CLUSTER SUFFICIENCY OF THE FAMILY

Sample	*N*	Mean	Variance
1	53	15.66	7.55
2	48	13.79	14.98
3	44	15.98	8.35
4	42	14.57	9.47

TABLE 36

ITEMS FOR THE CONCEPT OCCUPATIONAL-EDUCATIONAL ACHIEVEMENT DEFINED BY ITEMS FROM THE INTERVIEW
(Significant group or sample differences reported for individual items)

Family Occupational-Educational Achievement Items	Point of Dichotomy 1	Point of Dichotomy 0	Group Differences
Father's occupation (37)	prof., skilled	semi-skilled	
Father's education (39)	H.S.	no H.S.	ND > D
Earned income (41)	steady	other	
Mother's occupation (46)	skilled or housewife	semi-skilled	4 lower than 3; D lower than ND
Mother's education (48)	H.S.	no H.S.	
Type of dwelling (58)	owned	rented	ND better than D
Adequacy of home (59)	adequate	inadequate	ND better than D

Note.—Numbers in parentheses refer to interview item number.

logy or from a theory of learning. It is appropriate to rebel against parent figures more than against society, and it is healthy to discriminate between the two classes of individuals instead of promiscuously treating them the same.

Information from Mother and Son Compared

In this section the overlapping items in the interview and History Record were compared. Two points, the amount of agreement (vs. distortion) in parents' vs. sons' reports and also the kind of distortion were investigated. It is necessary to recognize at the outset that the truth is not known since the informants were not further checked upon.

TABLE 37

ANALYSIS OF VARIANCE FOR THE CLUSTER FAMILY OCCUPATIONAL-EDUCATIONAL ACHIEVEMENT

Source of Variation	*df*	Variance	*F*
Between samples	3	12.73	
Code	1	1.53	
Delinquency	1	36.63	12.02**
Code × Delinquency	1	.04	
Within samples	183	3.048	
Total	186	596.093	

** $P < .01$

TABLE 38

SAMPLE MEANS AND VARIANCES FOR CLUSTER FAMILY OCCUPATIONAL-EDUCATIONAL ACHIEVEMENT

Sample	*N*	Mean	Variance
1	53	5.04	2.88
2	48	4.13	3.47
3	44	5.19	2.32
4	42	4.33	3.11

Procedure

The Samples. The number of cases in each sample with both an interview and a History Record is given in Table 13 ($N = 143$). There was no significant difference in the number of cases with both interview and History Record either when the code groups or the delinquent-nondelinquent groups were compared.

The Items. The total interview consisted of 157 items, and the History Record had 138 items. The total number of items that could be analyzed with respect to the agreement or disagreement was 38.

Scoring. The identical or equivalent items in the two forms were multiple-choice type, and were designed to provide one appropriate response for each member of the population sampled. For some items the response alternatives were arranged in order from "bad" and "good." That is, some responses could be considered more socially acceptable or more "normal" in general than others. Twenty-three of the 38 items included in the study permitted a directional type of scoring.

First, the 23 items were considered as a set on which there might have been a tendency to distortion of answers by either the *S*s or their mothers. The items showing disagreement in each case were totaled separately for the number of items on which the mothers' account or the sons' account suggested better adjustment.

Secondly, the samples were compared according to the number of agreements or disagreements on the total set of 38 items. From these scores, two groups were selected, one with high agreement and one with low agreement scores. These groups were then compared according to the score obtained on the clusters of items discussed above.

Results

Reliability. Out of a possible 5434 agreements in the 38 item set, there were 3340, which is 61%. Out of a possible 3289 agreements on the 23 item set, there were 1943 agreements. In addition, there were 1040, or 32%, of the answers that were only one-step differences, in the 23-item set. Thus 91% of the answers in this set were either

TABLE 39

MATRICES OF CORRELATION COEFFICIENTS FOR FIVE CLUSTERS OF ITEMS PRESENTED FOR EACH OF FOUR SAMPLES

Cluster	Family Sufficiency	Occupation Education	Achievement	Psychopathy	Conflict
Sample 1					
Family sufficiency	—	.53*	.26	−.33*	−.38*
Occ.-Ed.		—	.55*	−.13	−.25
Achievement			—	−.22	−.22
Psychopathy				—	.72*
Conflict					—
Sample 2					
Family sufficiency	—	.30	.02	−.31	.13
Occ.-Ed.		—	.30	−.18	−.25
Achievement			—	−.25	−.30
Psychopathy				—	.26
Conflict					—
Sample 3					
Family sufficiency	—	.45*	.17	−.24	−.26
Occ.-Ed.		—	.32*	.05	−.26
Achievement			—	−.34*	−.08
Psychopathy				—	.53*
Conflict					—
Sample 4					
Family sufficiency	—	.20	.08	.09	−.03
Occ.-Ed.		—	.32	.14	−.06
Achievement			—	−.11	−.07
Psychopathy				—	.55*
Conflict					—

* Significant at the .05 level or less.

agreements or one-step disagreements. The remaining 9% of the answers were either unanswered or showed large differences.

The Twenty-Three Item Set. Table 40 contains the results of *t* tests for the differences between samples or groups on these items. These results show no differences among the samples in the degree to which the sons distort. However, the mothers of the Sample 4 boys had higher disagreement scores than other groups, indicating a bettering of the picture by those parents. This is reflected in greater distortion by Sample 4 mothers compared with Sample 2 parents, with Sample 3 parents, and in combined delinquency samples and code samples.

The Thirty-Eight Item Set. Differences between samples or groups on the 38-item set were nearly the same as for the 23-item set.

The High- and Low-Agreement Groups. From the distribution of agreement scores for the total group, two subgroups were drawn, those with high agreement scores and those with low agreement scores on the 38-item set. Twenty-seven percent of the total *N* of 143 was in each group. Those in the high group agreed on 26 or more of the 38 items, those in the low group on 20 or less of the 38 items. One individual with 21 agreements was included to make both groups consist of 39 members. A test of independence of the two variables, code and delinquency, suggested that the low-agreement group contained more delinquents, and the high-agreement group had more Ss from Sample 1 than would be expected ($P < .01$).

TABLE 40

SIGNIFICANCE TEST RESULTS (*t*) FOR THE SUM OF ITEM DISAGREEMENTS INDICATING A MORE FAVORABLE PRESENTATION OF THE HISTORY BY EITHER THE SON OR THE MOTHER[a]

Comparison Sample or Group:	Mother More Favorable	Son More Favorable	Total Agreement
2 vs. 4	.01 (4)	n.s.	.01 (2)
3 vs. 4	.01 (4)	n.s.	.01 (3)
1 vs. 2	n.s.	n.s.	n.s.
D vs. ND	.01 (D)	n.s.	.01 (ND)
025 vs. 489	.01 (489)	n.s.	.01 (025)

[a] Indicated in parentheses is the sample or group with the higher mean.

In order to determine whether differences in the high- and low-agreement groups on the five clusters would remain when each sample was considered separately, high- and low-agreement groups were taken from each of the four samples. Within-sample comparisons were then made; that is, the high-agreement group from Sample 1 was compared with the low-agreement group from Sample. 2, and similarly with the other samples. When this was done, differences between means appeared in Sample 3, where clusters of *Psychopathy, Family Sufficiency* ($P < .01$), and *Family Occupational Level* ($P < .05$), showed reliable differences. The high-agreement cases were low in *Psychopathy* and high in the other two clusters.

DISCUSSION

Methodology

In this section certain points that relate to the methodology and to the concepts under study are discussed. The samples were intended to represent four populations. The four populations are natural, as opposed to artificial groups, and might, but for changes in the culture, be reproduced at any time. It is important to note that neither MMPI code nor delinquency status implies as does, say, psychodiagnosis, an interpopulation homogeneity. These phenotypical characteristics of the populations are interesting for a priori reasons.

It was noted that at the ninth-grade level certain code or test results did absorb some of the variance in the subsequent behavior of the testees. Within each of the code populations (groups), over 20% of the entire ninth-grade population was included. The two groups contributed differentially to the number of delinquent boys. That code should be a predictor of delinquency in this limited sense suggests rather than confirms the idea that personality factors are influential in the occurrence of delinquency. (It can be pointed out that not too long ago this contention would have been considered far from satisfactory, i.e., of little utility.)

Delinquency Rating

We followed Hathaway and Monachesi in defining delinquency in terms of crimes recorded by the police. Such a definition provides a categorical classification system of delinquent and nondelinquent. Only the facts of the individual cases are subject to question. The severity of delinquency as rated here, or as rated in the Hathaway-Monachesi study, requires more interpretation.

On the whole, the pattern presented in the two delinquent samples suggested that delinquency is an adolescent phenomenon. The earliest period records (before the *S*s reached age 17) had the highest incidence of delinquency. Each of the two subsequent follow-up periods revealed a marked drop in the percentage of samples in contact with the law (83%, 60%, 21%). Thus, delinquency appears to be concentrated in the years between ages 14 and 19.

The ratings of each record for the period in question were based upon reliable expert judgment. This is their greatest strength. The statistical properties of the combined ratings for the different age periods gave no hint that there was any continuity in behavior from period to period. From a rational point of view it is conceivable that delinquency may be a scaleable variable when the social significance and desirability of the acts are under consideration. That such a scale should also be a natural dimension of personality with which behavior is to be understood is, of course, unproven if not unlikely. The rating used in this study to give an a priori estimate of the "psychological delinquency" involved in each record, was a rating of the entire period under consideration. A highly reliable nine-point scale was used.

As a digression, it is worthwhile to consider the meaning of results limited by a foundation as imperfect as the present one. No one with scientific sophistication has ever supposed that delinquency represents a major variable in the understanding of man. This is not to say that other more general concepts will not be reflected differentially in delinquent and nondelinquent samples, or that this is not important. It can only be hoped that the more essential characteristics of delinquent and nondelinquent samples are both psychologically and socially meaningful. We must be prepared for the possibility that no immediate psychological integration can be made of the factors that are found to be associated with or characteristic of the delinquent. An obvious example of this kind of explanation may be found in the purported (psychological) background of the delinquent wherein economic deprivation, broken home, etc. are noted. Most psychologists will allow that these factors often have psychological import. That each situation may be reduced to a psychodynamic influence without numerous gratuitous assumptions is, of course, impossible. This seems to indicate that the translation of factors in the sociological or economic language systems into the psychological system must for the present be accepted as incomplete.

Thus, it appears that the concept of severity of delinquency is probably more a dimension that is used by the expert observer to describe society's reaction to certain acts than a dimension that contributes to understanding the delinquent. Similarly, the factors related to delinquency will often be less psychological than economic or sociological.

Sampling Adequacy

Sampling difficulties altered the representative qualities of the data gathered through the interview and the History Record. If judgment is based on the Hathaway-Monachesi ratings made in 1952, the unsampled cases in both the interview and the History Record subsamples were poorer achievers in high school, less educated, and used their time badly when out of school. In addition, the History Record sample, which depended upon a parental informant, had fewer cases of broken homes. The interview sample missed significantly more of the younger members, presumably because they were in service.

The social agency follow-up data showed that the cases without the interview and History Record had significantly more instances of *Family Disruption Due to Disease* and *Poverty or Economic Need.* In addition, the History Record sample contained less *Dissocial Behavior* and *Psychiatry for the Family* than would be expected. It was among the delinquents, and especially among the 489 delinquents, that these differences occurred. These comparisons suggested that the grossly dependent families were not interviewed or questioned adequately. Thus, sampling deficiencies clearly reduced the contrast between code types and delinquent types.

The inadequacies of sampling had no appreciable influence on the mean delinquency rating. This presents something of a paradox, for in the social agency data, the severe delinquents had more points on a number of categories than did the mild delinquents. Since some of the same categories reflected differences in sampling among the interviewed or History Record cases, presumably the mean delinquency rating should have been less in these subgroups. Apparently the group without agency contact also contained people with high delinquency ratings, and these cases tended to be interviewed; those in the group with agency contact who accumulated many points, and had high delinquency ratings, tended not to be interviewed (or to have History Records).

Item Reliability and Validity

This study dealt with data which were recorded in a standard format and thus were in item form. Each item is either true, true to some extent (as in the *Q* sort), or not true within the limits of our knowledge.

The items were examined individually, and, in a few cases, in groups or scales. In all instances, the presumed meaning of the item (or category or scale) was its obvious semantic reduction. This seems to be a justifiable approach to some types of data, and a rather naive one for others. If the narrowly factual items are dismissed from question, a fairly large number remain that may or may not represent some past or present state of affairs. Further, item variance may be determined by other states of affairs than are connoted.

In reviewing the areas covered, the social agency data seem to be most certainly what they purport to be. That is, notations of a certain type are in the agency records. In the History Record and interview, some recourse may be made to the study of duplicated items. Even within this set it must be recognized that no actual validation of the data used is possible. That is, the results did not represent the objectively true state of affairs, but rather the point of view of those supplying answers to the items on the History Record and the interview. Nor was there any overall measure of scale reliability in the sense of consistency of responses when the same person was measured twice, or the halves of a test compared.

From the number of agreements between answers to comparable questions on the interview and the History Record, there is some indication of a type of reliability for the measures. Only 38 items, approximately one-fourth of the items used, were similar enough to be compared. There were 143 *S*s with both an interview and a History Record. This was not a random or representative set of items or *S*s. Sixty-one percent of the responses on the 38 items were in exact agreement. All but 9% of answers on the History Record and the interview were within one point. Considering the fact that the questions were multiple choice, with several possible responses to each, the reliability, measured in this way, was fairly high.

The concept of reliability in connection with agreement or disagreement as measured here seems to be a rather subtle one, dependent upon relationships within the family, and of the family with society. Still, agreement does constitute a type of reliability, in that the two judgments are the same. Perhaps there is even some slight indication of validity, if it can be said that concurring judgments probably have some basis in fact. The facts of each case often are less important than their interpretation. These appear to be reliably measured in the interview and the History Record.

Implications for Community Action

The implications of our study which are of methodological interest have been discussed. Because of the enormity of the social import of delinquent behavior, some comment on those conclusions from this study which have relevance for prevention and treatment should be made.

Public health efforts to control the occurrence of disease have always met with greatest success when they were directed toward prevention rather than toward treatment. The same general philosophy has usually been expressed by those interested in the control of psychological disorders, and of social, economic, and political ills. The present efforts of most communities, however, are largely concerned with the treatment of individuals already in social difficulty. This is necessarily the case when there is a large number of delinquents and too few policemen, probation officers, agency social workers, and almost no other professional personnel to deal with the individuals requiring current attention.

Sociological studies and the experience of almost everyone who has worked in a community agency show that there is a convergent pattern in which disease, poverty, mental illness, and immorality in the family environment are all part of the same cluster to which crime also belongs. Our study confirms this finding.

In general we found a substantial coincidence of social agency contact by family members and later delinquency in the sons. Sometimes this need for community support of the family was expressed even before the birth of the to-be-delinquent child. For the most part, agency activity in the period of interest here consisted of rather poorly integrated, fragmentary efforts by various agencies. Each agency was mainly concerned with its unique role: relief, hospitalization, probation, psychotherapy, or whatever. Little really concerted effort to treat the whole array of disorders within the family was then feasible. Both our study and the practical experience of most social workers would suggest that the more recent efforts of agencies and the more uniform training of social workers will contribute to improving coordinated activities of community resources. Stevenson (1956) has stressed the same point in his suggestions for the planning of mental health programs.

Social agencies, schools, hospitals, clinics—and the public generally—have been profoundly influenced by the theories and practice of dynamic psychiatry. The impact of psychoanalytic theory on social work training is substantial. In our view, this theoretical orientation has in some cases been misguided. It is, of course, perfectly reasonable that the insights of new and powerful psychological theory should be brought to bear on problems of social discord; effort in this direction has had undoubted value. Social casework in the area of delinquency is not often strictly individually oriented psychotherapy. The psychotherapy that has been done, however, has not met very much success. We believe that one reason for this is a failure to select properly those cases most likely to benefit from treatment.

The data show that delinquents were described as persons who "act out, are nonconforming, extrapunitive, unpredictable, self-indulgent, envious, deceitful, critical, sensitive to demands, and give up when frustrated." These are not easily modified traits. They are considered expressions of disordered character structure rather than symptoms of neurosis. Psychotherapeutic intervention as a treatment method is rarely successful with such persons. On the other hand, those cases in which delinquent acts are expressions of neurotic conflict are generally considered to have a better prognosis. Our estimate of the proportion of individuals in the general population whose delinquency may be an overt expression of neurosis is something less than 7%. It would seem that selection from the neurotic cases would increase the proportion of improved cases among those few who can be treated psychotherapeutically. Furthermore, it was found in this study that it is possible to detect differences in personality in early adolescence before delinquent behavior is exhibited. This means that case findings could serve the cause of prevention by locating individuals who are both potential offenders and good therapeutic prospects. The implementation of such case finding would require evidence from both personality testing and from the family social history. To do this would, obviously, involve expansion of community agency activities and staff. But even without redirecting psychotherapeutic efforts, the information we have described can be of benefit.

The larger problem is in finding preventive methods appropriate for the biggest group of delinquents. In this study that group was represented by the boys in Sample 4. A number of social forces have been brought to bear on this problem in the past. These derive both from common sense and from results of sociological surveys. These show, as our data confirm, that the social roots of delinquency are in marginal areas of the community where poverty, disease, and bad example are common and where both subcultural and family environments are conducive to delinquent behavior. In order to overcome these deleterious influences, communities have initiated such programs as housing redevelopment, improved schools and playgrounds, youth centers, and organizations sponsored by both religious and community groups, relief agencies for the sick and the poor, and special courts and services for juvenile offenders.

All of these appear to be reasonable approaches to the problem but they have not been as effective as hoped. As a consequence there has been a tendency for the public, the press, and even some officials to express alarm at what appears to be an increasing crime rate among juvenile offenders. The reasons for this apparent increase are not particularly germane to the purposes of this discussion, but among them we may include the following. There have been rapid technological changes in our society which have resulted in urbanization of the population. Thus, there is less opportunity for freedom of movement or of individual expression without interference with the rights of others. The greater use of the automobile is a real danger to others, since youth characteristically seeks it for excitement and adventure. In addition, improved methods of mass communication, law enforcement, and actuarial accuracy emphasize the problem. If, for example, traffic offenses are excluded from the statistics and if they are reported as rates (i.e. corrected for the enormous growth in population), modern youth is probably not much more delinquent than were previous generations.

It should be noted that the use of large scale preventive programs is a relatively new development. While there is considerable hue and cry about the failure of some of these approaches—and it is pointed out that twenty years of social reform have only increased the crime rate—the fact is that the impact of such prophylactic approaches cannot be assessed more quickly than the lifetime of one generation. It requires the growth of a new generation, to whom the preventive measures are presumably being applied, to show the effects of any such programs. And such assessment must account for social change which confounds the results of these programs.

The average age of offenders has steadily declined in recent years. The contemporary pattern of separate juvenile courts, probation services, and reformatories are reflections of the recognition both by the public and by the legal and social work professions that the situation requires modification of techniques designed for the treatment of adult offenders. A clear result of these modern practices has been a tendency to substitute probation for confinement. This mirrors a change in attitude from one of punishment for crime to one of rehabilitation of the criminal.

That it is the juvenile delinquent who grows up to be the adult criminal is well established. Very few people commit their first offense at an age beyond the earliest adult years. But there are relatively few adult criminals and a fair number of juvenile delinquents. The present study shows that whatever complex of environmental and personality factors may be influential in the expression of delinquency, it is more than anything else a developmental phenomenon of adolescence. The rate of delinquency declines markedly with increasing maturity after age 19. All of this suggests that there is value in the current practice of liberal, supervised probation. Much notoriety is given to teen-age crime and pressure to deal harshly with the offender may be defended if the goal is prompt punishment. But the suggestion that current juvenile court practices contribute to recidivism cannot be defended. Indeed, we believe the opposite interpretation to be correct.

With conditions extant in our changing society and with the prospect of continued population growth, efforts at treatment are almost surely doomed. The only hope for a substantial reduction in the number of offenders is to look for methods of prevention while trying to improve current methods of treatment. As was pointed out, the well-considered community improvement programs have been disappointing. The probability is that these programs have been of great value to those who use them, but that these do not include a sufficient number of the delinquency-prone individuals. Again, therefore, we recommend case finding. Prevention programs, however well conceived, will not work if self-selection operates. They may work if participation in them by the potential delinquents is insured. Following from the work of Hathaway and Monachesi, the findings in this study clearly demonstrate a high association between certain personality characteristics and a tendency to delinquency. The testing also shows a group of personal characteristics which are associated with nondelinquency. The study has shown that regardless of personality characteristics, individuals reared in socially inadequate homes and community environments have a high rate of delinquency. At the same time, the individual whose milieu would be considered socially acceptable is unlikely to become delinquent despite his personality characteristics. Nearly half of the boys become delinquent when their personality traits are of the delinquency excitor sort *and* their milieu has been of a destructive kind. The coincidence of these factors in the general population is substantial and it is with this delinquency-prone group that preventive measures might best be directed. At present the community efforts are unfocused and spread thin and, as was said, are largely dependent upon self-selection. Our recommendation is that these efforts be concentrated upon the groups we know to represent the highest concentration of potential offenders. Even if totally successful, neither case finding of this sort nor that which we recommended for the neurotic delinquent could be a complete solution to the social tragedy of delinquency. These methods still leave a very substantial number of factors and individuals unaccounted for. What the methods do provide are sound public health measures directed at the places most likely to give the best results.

Most delinquents do not appear to be motivated by special desires or needs for antisocial acts, but rather are only incidentally delinquent. We have found that these boys are superficial, sensual, and selfish. Their relations with others are casual and their chief interests are self-indulgent and characterized by needs for excitement and changes. They achieve poorly in terms of the social mores. These characteristics need not in themselves lead to delinquent behavior. For example, Sample 3 had personal traits similar to those of the delinquents, but did not become delinquent. However, Sample 3 for the most part came from very different homes than Sample 4. It is the combination of family history and defects of character which leads to the largest evidence of delinquency.

We agree with the conclusion of Hathaway and Monachesi that these youngsters need redirection into socially acceptable activities. Some efforts of this general type have been made by social workers in some large cities. Their clinical reports are encouraging. The approach requires that individuals (or groups) with a high probability of delinquency are sought out before the occurrence of delinquent acts requiring judicial action. This means that the personnel and facilities provided by the community should be designed in part to curb delinquency. Past efforts by communities, such as playgrounds and fellowship clubs, are places and organizations which do not attract the boys who "need" them. They fail, therefore, as adequate methods of treatment or of prevention. The facilities (and boys' group organizations) should be of a sort that provide opportunity for excitement without expecting very much in the way of achievement, personal loyalty, or responsibility. As an example of this sort of enterprise we cite those communities which support a group of "hot rod" enthusiasts who, under the leadership of responsible adults (e.g., police or probation officers), are led into interest in traffic safety in addition to racing around a "drag track" provided by the citizens and learning about ways to "hop up" a gasoline engine.

If our social agencies are to be the avenue of approach to the delinquent, through detection of such individuals (and their families) at an early age, treatment programs for them must be redesigned. To do so first requires the employment of workers whose personalities and personal values are such as to make this work possible for them. If one's personal values are such as to abhor indifference, lack of achievement, promiscuity, and superficiality of these boys and their families, then little can be expected. The workers themselves probably ought to be people who are energetic, aggressive, and enjoy thrills and adventure.

This study contributes some tangential data on questions concerning personality development. Johnson and her co-workers (Johnson & Szurek, 1952) have long argued the thesis that the delinquent acts out the unconscious desires of his parents, particularly the destructive impulses of his mother. Some of the trends we find showing differences between the reports of mothers of delinquents and those of nondelinquents lend some support to Johnson's theory. The mothers of the delinquents are more generous in evaluating the home and family success than are the mothers of the nondelinquents. This judgment is based upon the assumption that the delinquent son will not be more critical than the nondelinquent son.

More than parental attitudes appears to be involved. The social and economic level of adjustment by the boys in the samples seems to be a joint function of their origins and their personalities. An 025 code predisposes the boy to achievement greater than that of his family, while 489 code boys are likely to underachieve their families. Low family social and economic level leads to delinquency, regardless of personality pattern. If society wishes to affect development it would require some substitute for the early family influences. Even though the boys in our Sample 2 went to work early, they became involved in delinquent acts. While they may, like their counterparts in Sample 1, have had some potential for mature functioning, their fate became entangled with the consequences of their delinquency and the poor home example that had been set for them. They did not achieve educationally or vocationally as much as their potential would suggest (though they were higher than their parents in these areas). It is possible that early preventive measures with these boys, which would allow them to escape from their homes, but would provide a stable environment instead of forcing too early efforts at independence, might have been effective.

The group which represents the largest segment of the general population is Sample 3. These boys are examples of the greatest waste of our most important national wealth: youth. While these young men do not cause serious trouble in the community, neither do they contribute to it Their general personality characteristics are similar to the largest group of delinquents. That is, they are carefree, impetuous, and little interested in cultural values. They underachieve their parents and their own potential. They lack productiveness and fail to contribute to the general welfare. If we consider it desirable to exploit the intellectual and moral potentials of the nation, then these people have been wasted. If, on the other hand, we take a more permissive philosophy of self-determinism within the bounds of law, the matter of economy and conservation of national resources is of less concern. The current tenor of the American Society, as contrasted with contemporary France, for example, it seems to us, is toward greater concern with the maximum utility of human resources. If this is so, society may wish to be concerned about those boys who represent a substantial proportion of persons their age. It seems likely that the kind of syndrome they exhibit by age 14 and after could be detected early. Little effort has been made in the way of parent education, community action, or school training methods to alter the course of development in these relatively normal children. As adults they seem fairly happy people and one might wish to question the right of others to impose value oriented influences on them as youngsters. But that philosophical dilemma is not a problem for this paper. From our own point of view, however, it can be pointed out that society already attempts to teach certain values in its schools, churches and other institutions, some of which are directly opposite in expression from those exhibited by these nondelinquent boys. An example of this tendency is found in our statistics on sexual behavior. These boys show much earlier, more frequent sexual activity than their counterparts in the conforming Sample 1. While these activities do not lead them to become labeled as delinquent by our definition, they are counter to the expressed American moral climate. Whether such behavior is to be accounted for by an hypothesis of greater drive or one of less inhibition cannot be determined in our study. It would be our guess that the latter hypothesis is the correct one, since this group shows generally diminished control functions. We do not know what society can or should do about these things, but it seems clear that many of the moral values of the culture are not followed by these boys. It is possible that they are badly taught by their parents. In any case, we believe these characteristics are formed early, certainly before adolescence, and are thereafter resistant to change. Whatever intervention with them or their families is attempted must, therefore, be early.

SUMMARY

Hathaway and Monachesi found that certain MMPI results from testing ninth graders were related to subsequent delinquency. Elevation on the Psychopathic Deviate, the Schizophrenia, and the Manic scales (code number 4, 8, and 9, respectively) were associated with high rates of delinquency while elevation on the Social Introversion, the Depression, and the Masculinity-Femininity Scales (code numbers 0, 2, and 5, respectively) were associated with a low rate of delinquency. If both of the two high-point scales were 4, 8, or 9 (i.e., 48, 84, 49, 94, 89, 98), the average delinquency rate was 33%. If one of the two high-point scales was 0, 2, or 5, the rate was 17%. The average rate for all other codes was 20%.

Delinquency was rated by Hathaway and Monachesi two years after testing (approximately age 16) and 4 years after testing (approximately age 18). Boys who were called delinquent had police and court records which revealed more severe delinquency than simple traffic offenses.

For this study, four samples of boys were drawn from the original population of 1958 boys tested by Hathaway and Monachesi in 1947-1948. Two samples, delinquent and nondelinquent, were drawn from the high delinquency code group (489) and two were drawn from the low delinquency code group (025). All case records were rechecked and brought up to date. Those who were delinquent before testing and not again were excluded as were cases only questionably delinquent. Cases misclassified in the 1950 and 1952 follow-ups, or whose classification of nondelinquent had changed to delinquent by 1956, were excluded. The following samples remained:

Sample 1: Nondelinquents with 025 codes as described above ($N = 71$) from a population of 393.

Sample 2: Delinquents with 025 codes ($N = 71$) from a population of 71.

Sample 3: Nondelinquents with 489 codes ($N = 71$) from a population of 274.

Sample 4: Delinquents with 489 codes ($N = 73$) from a population of 129.

Although the occurrence of delinquency was nearly two times as great among boys with 489 codes, the severity of delinquency was not related to this difference, for the mean ratings of the 025 delinquent cases nearly equalled the mean ratings of the 489 cases. The rate of delinquency for the two delinquent samples was found to be 82% for the first period (through the 16th year), 60% for the second period (17 through 19 years) and 20% for the third period (20 through 23 years). Ratings given the delinquent cases during the three periods failed to show a reliable correlation from period to period.

Three individual approaches were made to study the histories and adjustment of the boys in the samples. The records from all voluntary social agencies were examined for information relating to the families of the boys. The boys themselves were interviewed where interviewing was possible. In the interview, historical, adjustment, and personality data were obtained, as well as a final personality description on a 115-item *Q* sort. The mothers of the boys were asked to complete a standard questionnaire concerning their sons.

The success of these three follow-up techniques varied from complete coverage of the samples in the social agency investigation to two-thirds coverage in the interview and 60% coverage in the information gained from the maternal informant. The cases that were missed tended to reduce the contrast between delinquent and nondelinquent samples. There was no suggestion that the results as obtained should be discounted. More exactly, the items not reflecting sample differences should not necessarily be counted as irrelevant to the sample characteristics.

Methodological problems arise in another area besides sampling. Item reliability, although not sampled uniformly, appears to be respectably high. This interpretation was made from the overlapping 38 items in the interview of the *S*s and the questionnaire sent the mother.

Further, the fact of agency contact by the family was noted and analyzed. The data gathered from the social agency records were organized into seven descriptive categories. They were *Family Disruption Due to Disease, Poverty or Economic Need, Dissocial Behavior in the Family, Psychiatry for the Family, Marital Disruption, Inadequate Parent-Child Relationship,* and *Minor Psychological Problems.* The interview data were first analyzed as items and then two clusters, *Family Sufficiency* and *Family Occupational-Educational Level,* were studied. Personality descriptions in the form of *Q*-sort ratings were compared. The results from the parent questionnaire were analyzed as individual items, and three clusters were formed. These clusters were *Achievement, Psychopathy,* and *Parent-Child Conflict.*

No reliable relationship between MMPI code type and families in contact with a social agency appeared. The two factors, MMPI code and agency contact, taken together discriminate groups with the following differential rates of delinquency. Forty-two percent for the group of boys who had 489 codes and whose families had contact with a social agency became delinquent; the group of 489 boys without such agency contact in their families was only 23% delinquent; 22% of the individuals with 025 codes and families with agency contact became delinquent, while the 025 boys coming from families with no agency contact were only 11% delinquent.

Below, the findings are discussed for code and delinquency classification. First the differentiating attributes of delinquency within the code samples will be considered.

Sample 1 vs. Sample 2. In their origins the samples differed widely; the education, occupation, and income of the parents of Sample 1 were higher than those of Sample 2. This is an understatement, for Sample 2 was characterized by the agency categories of disease, poverty, and dissocial behavior, which was almost absent in Sample 1. The homes from which the boys in Sample 2 came were less sufficient, in both physical and moral sense, than the homes of any other sample. With regard to the boys, there was more parent-child conflict, jealousy of siblings, and fear of the parents in Sample 2. Sample 2 appears to have had more conflict with their parents than any other sample.

In their own personality characteristics, the *S*s of Sample 2 were differentiated from those of Sample 1 by valuing humor and entertainment, and day dreaming more. They were said to be more responsive to their own needs and impulses, to be less moralistic, strict, and sarcastic, but more masculine. They were more psychopathic than all other samples. Their performance in school was trouble-ridden and poor, as was their general level of achievement.

As to the characteristics of their social adjustment, Sample 2 contrasted to Sample 1, left home early, had less residual of attachment to their homes, and usually carried on their activities away from home. Their friends were from a lower social and economic standing and were more often delinquent and more often married than the friends of Sample 1. In essence, then, these boys were alike in that they made an adjustment that was more like that of their parents than like each other. They differed in that Sample 2 rejected the parental home and Sample 1 did not.

Sample 3 vs. Sample 4. There were relatively fewer items differentiating these samples than the 025 samples. This may be due to the sampling inadequacies noted above, which reduced the contrast between these two samples.

In background, the clusters *Family Sufficiency* and *Family Occupational-Educational Level* reflected especially great differences between the means of Samples 3 and 4. Sample 3 was the highest (best) on both of these measures within the four samples. (Sample 4 was higher on both of these clusters than Sample 2.) In the data gained from social agencies, there was a record of disease, poverty, and dissocial behavior associated with the family from Sample 4. In addition, psychological problems appeared more frequently among the delinquent cases. In their family life, the delinquent sample had mothers who were away more, and they associated with their siblings less frequently. The boys and their mothers viewed things differently at the time of the study. In Sample 4, the mothers were prone to answer questions about the family history as though things had been better than their sons saw them.

As to the boys themselves, the items differentiating the two 489 samples referred to delinquent-like behavior in Sample 4. There was more sexual promiscuity, conflict at school, and fights. Sample 3 was said to be more orderly, intellectual and submissive, while Sample 4 was considered deceitful, critical, nonconforming, and inclined to act out. Sample 4 was also said to be sensitive to demands, to give up when frustrated, and to have more sexual activity.

In achievement, while the two samples did not differ greatly, Sample 3 was somewhat better. Although there was not a great difference between psychopathy scores for the two samples, Sample 3 was less psychopathic than Sample 4. The only recorded difference in the social level taken by the boys in these two samples was the differing number of delinquents among their friends. The delinquent sample had more delinquent friends.

In general, the delinquent sample came from much worse homes than the nondelinquent sample. The differences in adjustment to life were less spectacular than the differences between the two 025 samples. Delinquency and background were the major differences between the two 489 samples. Present performance, level of friends, and adherence to family ties reflected no difference between samples as they did in the 025 samples.

Delinquency vs. Nondelinquency. This comparison reiterates the differences mentioned above. It does not seem enlightening to repeat these results.

Code Comparisons. The code groups did not differ greatly. Within the nondelinquent samples (1 vs. 3) the differences reflected the rather conservative, high-achievement records of Sample 1 as contrasted to the adventurous, extroverted, but low achievement of Sample 3. Within the delinquent samples (2 vs. 4) Sample 2 had less trouble (in school and due to drink) while they had more married friends and less contact with churches.

The over-all comparison of the code groups suggested more intrapsychic difficulties in the families of the 489 groups (dissocial behavior, psychiatric contact, and psychological problems). The 489 group was also said to be punished less often. The 489 group had easier sibling relations and tended to center activities away from home.

The amount of agreement between mothers and sons was less in the 489 samples. The mothers of the boys tended to think of their family history as better than did the sons. Thus, it may be said that the family histories of boys with 489 profiles, with respect to their own judgment, were more generously evaluated by their mothers, while boys with 025 codes make a more generous evaluation than their mothers.

The results were interpreted as having some bearing on theories of personality development and of delinquency. First, delinquency is for the most part an adolescent phenomenon. We believe that the characteristics expressed in the personalities of the boys studied develop very early and, in some cases, they may be constitutional derivatives. A deleterious environment in itself is not a sufficient explanation for certain expressions of personality, but on the other hand, such an environment does show a high degree of association with delinquency as an overt behavior. There is some evidence that bad example on the part of the parents may in subtle ways instill the learning that results in subsequent delinquent behavior.

In terms of social actions and consequences for community living, more case finding and increased financial support of prevention programs were recommended. Careful diagnostic study together with careful evaluation of historical data might yield a small, but significant number of potential delinquents who could benefit from traditional forms of psychotherapy. For most cases the solution appears to lie in case management and in prevention rather than treatment. Current practices of juvenile court procedures appeared to be helpful in prevention of adult crime. Crime prevention programs for the modal group of predelinquent children should also involve case finding. The delinquency prone youngsters would be encouraged to participate in programs designed to fit their interests and personalities and be led by individuals equipped by temperament and training to work with such groups. It was pointed out that any prevention program requires a generation in time to evaluate the results.

Finally, it was suggested that there is considerable waste of personal resources in a large group of these boys even though they do not become delinquent. The alternatives open to the community in appraising this finding were discussed.

REFERENCES

Block, J. *The IPAR Clinical Q-set (Form II): 115 personality-descriptive items.* Berkeley: Institute of Personality Assessment and Research, 1954.

Briggs, P. F. Preliminary validation of a standard personal history for psychiatric diagnosis. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.

Briggs, P. F. Eight item clusters for use with the M-B History Record. *J. clin. Psychol.*, in press.

Cleckley, H. *The mask of sanity.* St. Louis: Mosby, 1941.

Glueck, S., & Glueck, Eleanor. *One thousand juvenile delinquents.* Cambridge: Harvard Univer. Press, 1934.

Glueck, S., & Glueck, Eleanor. *Juvenile delinquents grown up.* New York: Commonwealth Fund, 1940.

Glueck, S., & Glueck, Eleanor. *Unraveling juvenile delinquency.* New York: Commonwealth Fund, 1950.

Hathaway, S. R., & Monachesi, E. D. *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. Minnesota Press, Press, 1953.

Hathaway, S. R., & Monachesi, E. D. The personalities of predelinquent boys. *J. crim. Law Criminol. pol. Sci.*, 1957, **48**, 149-163.

Hoyt, C. J. Test reliability estimated by analysis of variance. *Psychometrika,* 1941, **6**, 153-160.

Johnson, Adelaide M., & Szurek, S. A. The genesis of antisocial acting out in children and adults. *Psychoanal. Quart.*, 1952, **21**, 323-343.

Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

Landers, B. *Towards an understanding of juvenile delinquency.* New York: Columbia Univer. Press, 1954.

Loevinger, Jane, Gleser, Goldine C., & DuBois, P. H. Maximizing the discriminating power of a multiple choice test. *Psychometrika,* 1953, **18**, 309-317.

Quay, H., & Peterson, D. R. A brief scale for juvenile delinquency. *J. clin. Psychol.*, **14**, 1958, 139-142.

Shaw, R., & McKay, H. D. *Juvenile delinquency and urban areas.* Chicago: Univer. Chicago Press, 1942.

Stevenson, G. S. *Mental health planning for social action.* New York: McGraw-Hill, 1956.

Sullinger, T. *Social determinants in juvenile delinquency.* New York: Wiley, 1936.

Wier, P. *Economic factors in Michigan delinquency.* New York: Columbia Univer. Press, 1944.

(Received December 2, 1958)

Vol. 73, No. 16 Whole No. 486, 1959

Psychological Monographs: General and Applied

IMITATIVE BEHAVIOR IN PRESCHOOL CHILDREN[1]

JOHN W. McDAVID

Iowa Child Welfare Research Station

THE CONCEPT of imitation is of direct interest to psychologists in several areas. Behavior which is patterned after a model set by some external figure constitutes a notable portion of human activity and operates as a critical mechanism in certain complex behavior. The phenomenon of social learning, possible in the human organism because of elaborate skills of communication, enables man to profit by observation of others around him and thus to short-circuit the long and tedious process of learning only through direct experience. This mechanism, especially in early childhood development, often derives from imitation. An older or more experienced person frequently serves as a guide for behavior before the imitating organism can "read" environmental cues for himself and behave independently in response to those cues. In later stages, the behavior of others serves to expose the imitator or direct his attention to relevant environmental cues.

Miller and Dollard (1941) have proposed an analysis of imitation and social learning by synthesizing elements of Hullian learning theory with empirical observations of human social behavior. A series of experimental laboratory studies, using both animals and children, served initially to document their theoretical analysis. Several more recent experiments (Church, 1957; Schein, 1954; Solomon & Coles, 1954; Wilson, 1958) have further explored the implications of their basic hypotheses.

The present study is designed to explore, within a framework comparable to that employed by Miller and Dollard, some potential sources of interindividual variance in the experimental learning of imitative behavior. In laboratory studies which assume all subjects to be equal and identical at the beginning of experimental manipulation, there is a certain component of variance ordinarily included in the error term which may be isolated as *personality variance.* The term applies to those variables which operate prior to experimentation to make subjects, in spite of traditional assumptions, *not* equal and *not* identical. Variables of this type may be conveniently categorized as (*a*) genetic or given biographical factors (e.g., age, sex, etc.) or (*b*) experiential factors or learned systems of habits (e.g., prior experienced exposure or specification of conditions). This research proposes to deal with both categories as sources of potential individual differences in behavior in the laboratory. The scope of the research reported here is broad, and the techniques utilized cover a wide range methodologically. The basic experiment may be considered a replication of the primary Miller and Dollard experiment on the learning of imitative behavior in young children (Miller & Dollard, 1941: Experiment V). These data, however, then constitute the basis for extended exploration of several factors: (*a*) biographical factors (age, sex of subject, and sex of leader), (*b*) intellectual factors (Binet IQ scores), (*c*) experiential variables (child rearing antecedents), (*d*) conditions of family structure (ordinal position in the family), and (*e*) tangential variables of social behavior which may gauge the generality of the imitative tendency over a wider area of behavior (observed dependency in play activity).

Although the data reported here are in many cases consistent with and supportive of research findings which have utilized gross and inclusive definitions of similar concepts (e.g., "dependency," "parental restrictiveness," or "social learning"), this study cannot serve as the basis for unguarded generalizations about gross variables of parent behavior, child behavior, or relationships between these. Certain of the concepts defined here may well be restricted to the particular operations of measurement or observation employed. A child who imitates adults as observed here may give evidence of some general characteristics in his day-to-day behavior, but one cannot conclude from the brief observation of a particular child in the laboratory that he is a "dependent" child as a teacher or another researcher might characterize him. The intent of this report is to describe a set of observations of a nature lacking in reported studies of imitation, both in terms of (*a*) individual differences in performance in the laboratory and (*b*) relationships between child rearing antecedents and observed imitative behavior in young children.

---

[1] The author is indebted to David H. Crowell for handling the administration of the Parental Practices Research Scale, to Ruth Updegraff, Rosalind Gold, and Nancy Mann for making available data from uncompleted M. A. research, and to June Naumann for her assistance as an experimenter and experimental leader.

THE BASIC EXPERIMENT

This study is intended to explore individual differences in the process of acquiring an "imitation habit." *Imitation* is defined as behavior which is patterned after a model or example provided by another organism. In contrast to "copying behavior," which Miller and Dollard (1941) define as a case of progressive modification of a complex response until it attains a particular criterion of identity to a model, the imitative behavior observed here is parallel to their definition of "matched dependent behavior," which they describe as a case in which the leader is able to "read" and to respond appropriately to relevant environmental cues while the follower is not. The responses are simple, and the follower depends upon the leader for a signal as to the behavioral act (already in his behavioral repertory) which is to be performed. Thus the concept of *social cue* arises. While the leader in this paradigm may be said to respond to an environmental cue (color, position, etc.), the follower responds to a social cue (the behavioral act of the leader) in determining his response. In function there is no difference, other than one of convenient categorization, between so-called environmental cues and social cues. A social cue, however, may be dealt with by the follower in two ways: it can either be handled as an end in itself (in which case no independent learning, in the sense of relating a behavioral response to a set of environmental stimulus conditions, occurs) or it may be handled as a means or instrument to an end of social learning (in which case the follower, after responding initially to the leader's act as a cue, eventually comes to "read" independently the same environmental cue to which the leader is responding and to behave independently under such conditions). Experimental studies with both animals (Church, 1957) and children (Miller & Dollard, 1941: Experiment X) have demonstrated transmission of learned behavior by means of social learning (the latter use of social cues described above). The experimental design employed here, however, like the initial Miller and Dollard experiments, precludes the occurrence of such social learning. These designs represent a limiting case in which the cues "read" by the leader are unavailable to the dependent subject. The individual disposed to the latter use of social cues should in these experiments suffer certain interference effects in learning the imitation habit, since in all cases the leader's response conveys invalid and useless information about the environmental cues which are present.

The basic laboratory study here involved a training series in which an experimental subject (the follower or dependent subject) was taught to imitate, that is, to make the same choice in a two-choice discrimination learning problem made by an adult (the leader) immediately preceding the subject. This tendency to make the same behavioral response made by the leader is referred to as the imitation habit. The experimental design parallels that of Miller and Dollard with two major exceptions. First, the follower was unable to observe whether the leader did or did not receive a reward for his choice in the problem. There were two reasons for introducing this departure. Such a procedure eliminates the "100% successful" prestige value which accrues upon observation of a consistently rewarded leader, thus increasing the latitude of individual differences in learning to imitate which should hypothetically relate to personality differences among the dependent subjects. In addition, the possibility of a higher order complex conceptual solution is eliminated. In the Miller and Dollard design, the experimenter is unable to state unequivocally that the dependent subject is rewarded for the act of imitation per se. Although in the animal studies (Church, 1957; Solomon & Coles, 1954) the rat may be assumed to be unable to observe whether the leader did or did not find food, in the experiments with children (Miller & Dollard, 1941: Experiment V; Wilson, 1958), the follower was aware of the consistent success of the leader. Under such conditions, it is conceivable that the follower, having had some experience with hiding-and-finding games, might derive a conceptual solution in learning to imitate, such as "Where the leader looks, there is where the experimenter is hiding the candy," or, in the case of learning to nonimitate, "If the leader found the candy, there is not likely to be any more left in that box." There is some evidence in the Miller and Dollard experiments that the second hypothesis may occur frequently, since they observed in first-trial behavior a preponderant disposition (77% of all responses) to respond in the fashion opposite to the leader's response. The concealment of success or failure of the leader from the follower should thus serve both to eliminate the possibility of such conceptual solutions and to emphasize individual differences in acquiring the imitative habit.

A second departure in design from the Miller and Dollard experiments involved the introduction of imposing color cues as irrelevant environmental cues. If the follower (for whatever reasons) is disposed to attend to environmental cues or to attend to the content of the information obtained in the observed leader's response, these irrelevant color cues should serve to interfere and retard the course of acquisition of the imitative habit. On the contrary, if the follower is disposed to attend to adults as cues for his behavior without regard to the informational content of their behavior, the irrelevant cues should have no effect upon learning the imitative habit.

In addition, while Miller and Dollard studied the imitative behavior of young children with both peer and adult leaders, this study is concerned only with the imitation of adult leaders.

Method

Subjects

Experimental *S*s for this study were 32 children between the ages of 45 and 67 mo., with a mean age of 54.6 mo. All *S*s were enrolled in the Iowa Child Welfare Research Station laboratory preschools, and had without exception previously participated as *S*s in other experimental studies. For the most part, this population derives from upper middle-class parents of professional standing who are relatively permanent residents of the community.

The experimental leaders were adults in their late twenties with whom the experimental *S*s were familiar by sight. In no case had a leader previously served as experimenter with any given *S*, nor had a leader ever been associated with a given *S* as a teacher.

Apparatus

The apparatus employed here was rather more complex than that of Miller and Dollard. In order to conceal from *S* information as to whether the leader was or was not rewarded, it was necessary to arrange some manner in which *S* could be informed of the leader's response without being informed of its consequences.

The problem box consisted of a small black wooden box containing two compartments (8″ × 8″ × 6″) whose front openings were covered by light wooden panels hinged at the top. The panel on one compartment was bright yellow and the other bright blue. On top of the problem box were two colored light bulbs, one above each compartment and of the color corresponding to the compartment door. The apparatus was arranged so that upon *S*'s response, the appropriate light was illuminated to record the choice made. The box was constructed with hinged panels on both the front and back of each compartment, allowing *E* to reverse the box and counterbalance for the position in which each color was presented. The problem box was placed on a low table at one end of a long rectangular room and, at the opposite end of the room (some 15 ft. removed from the problem box and isolated behind a wooden screen 5 ft. high and 6 ft. wide), two low chairs were placed facing a similar low table containing the report box. This consisted of a wooden box identical in every respect to the problem box, except that the colored door panels were not hinged. Lights above each compartment reported information in the same manner concerning the choice made at the problem box.

The reward for making the appropriate choice in responding was a single small candy-covered chocolate pellet which *S* was informed he could either eat immediately or save in a small sack provided by *E*. The candy was placed by *E* in a small round hollow in the center of the appropriate compartment on each trial.

Procedure

The experimental training session consisted of 24 trials. Half the *S*s were run in imitation of a female (in which case *E* was a male) and the other half in imitation of a male leader (in which case *E* was a female). The following instructions were read to *S* and the leader by *E*:

> *S* (subject's name), I would like for you to play a little game with *L* (leader's name) and me. You see, this little box (pointing to problem box) has two doors, a blue one and a yellow one. I'm going to hide a candy behind one of these doors, and I want you to guess which door it is behind. All you have to do is reach into whichever door you think the candy is behind and see if it is there. See? (*E* demonstrates that doors are hinged and shows *S* where candy is placed.) If you reach into the yellow door, then this little yellow light comes on (demonstrates) and if you reach into the blue door, this blue light comes on. If you like you may eat your candy when you find it. Or if you prefer, I will give you a sack to save it in. Which would you rather do, *L*? (*L* always chooses to eat candy immediately.) Which would you like to do, *S*?
>
> Now we're going to play the game like this: you will take turns looking for the candy. First *L* will get a turn, and then you will get a turn, *S*. Come over here and I will show you where to sit between turns while we are playing. (*E* directs *S* and *L* to chairs at opposite end of room.) You will sit on these chairs, and when it is *L*'s turn, I will call him (her); and when it is your turn, *S*, I will call you. There is a little box here just like the one I showed you where I am going to hide the candy. (*E* points to report box.) Only this one is just make-believe; it doesn't have real doors. Now you watch this box while *L* takes his (her) turn, and he (she) will watch it while you take your turn. It is fixed so that you can tell which box the other person looked into. If the little light on this side (pointing to yellow bulb) comes on, that means *L* thought the candy was behind the yellow door and reached in there. And if the little light over here (pointing to blue bulb) comes on, that means he (she) reached into the blue door. *L*, you will do the same thing while *S* takes his (her) turn, so you will know where he (she) thought the candy was. After you have taken your turn, you will go back and sit down and watch this box while the other person goes to take his turn. But don't ever tell each other whether or not you found the candy; that is a secret until the game is over.
>
> Do you understand how to play the game? O.K., now you sit here and I will go hide the candy and I'll call you when I am ready. (*E* returns to problem box.)
>
> O.K., *L*. You come and take your first turn. (*L* takes Trial #1.)
>
> O.K. Now, *S*, you come and take your turn. (*S* takes Trial #1.)

In pretesting, one practice trial was given each *S* before the beginning of the game, but this procedure was found to be unnecessary. The verbal instructions, although intricate, seemed adequate for explanation of the procedure.

S was 100% rewarded for imitation of the leader's response. The irrelevant color cues and position cues were counterbalanced to be 50% reinforced in a random order, such that in any block of four trials, *L* chose blue twice and yellow twice, each color once on the right and once on the left. The experimental session consisted of 24 trials (six blocks of four trials) in which *S* was preceded by the adult leader. Although approximately 44% of the *S*s had not yet learned the imitative solution to the problem during this time, it was discovered in pretesting that extension of the training series beyond this level was not likely to provide additional reliable data. In general, the problem was a relatively difficult one for preschoolers. Although motivation was fairly high at the outset, *S*s who reached Trial #20 with no more than chance success gave clear evidence of frustration and reduced interest in the task. It is quite possible that increased incentive could have been employed to counteract this. *S*s employed here had had some experience participating in various discrimination learning studies in which they are customarily rewarded with a relatively highly desired toy. The use of small amounts of candy as the sole reward obviously functioned less effectively as an incentive to sustained effort to solve the problem. Furthermore, the 24-trial training series required approximately 25 min., which approaches the child's tolerance limits for removal from preschool play.

For *S*s who never learned the imitative solution, the experimental series consisted of 24 trials followed by a termination trial; however, for *S*s who learned to imitate quickly, the series was abbreviated. The criterion for cessation of training was five consecutive rewarded (imitative) responses, following which a termination trial was added. The termination trial in all cases consisted of hiding candy in both compartments so that *S* was rewarded on his final trial regardless of his choice.

Although it was initially planned that the study would employ some 60 children, the fact that the experimental problem was relatively difficult greatly reduced the number of available *S*s. Pretesting indicated that children aged less than 42 mo. had considerable difficulty with the task. Of four pretest *S*s younger than this, three were unable to complete more than three trials before it became evident that they were either unable to comprehend and follow the instructions or unable to understand and utilize the information conveyed by the report box. For this reason, all *S*s aged less than 45 mo. were eliminated, reducing the total sample to 32 *S*s. In two cases, experimental irregularities made it necessary to replace *S*s: one *S* choked while eating his candy reward, and a second *S* momentarily lost interest in the problem and asked *E* to hold him on his lap.

The 32 *S*s were divided into equal subgroups on the basis of (*a*) sex, (*b*) age, split at the median of 56 mo., and (*c*) sex of leader to be imitated. Hence there were four *S*s in each of eight subgroups. For purposes of abbreviation, these subgroups are referred to by a three-letter code: the first letter denoting sex of *S*; the second, age of *S*; and the third, sex of leader. The subgroups, then, were: FYM, FYF, FOM, FOF, MYM, MYF, MOM, MOF.

Only one experimental observation was recorded on each trial: the occurrence of an imitative (rewarded) response. From these observations, three experimental measures were derived: (*a*) initial tendency to imitate, or the occurrence of an imitative response on Trial #1; (*b*) total imitative responses; and (*c*) deviation from chance imitation. Because the data were handled in blocks of four trials (each block containing counterbalanced color and position cues), *S* could score from 0 to 4 imitative responses within any given block. In the case of those *S*s who attained a criterion of five consecutive imitative responses prior to completion of all 24 trials, training was discontinued and the number of imitative responses was projected over 24 trials under the assumption that *S* would have continued to make consistent imitative responses had his training been continued. Therefore the imitation score (*b*) could range from 0 to 4 on any given trial block, or the total imitation score could range from 0 to 24. The deviation from chance imitation score (*c*) was derived on the basis of the assumption that chance would result in *S*'s making two imitative responses out of a block of four trials. The problem afforded only two alternative responses, and both color and position cues coincided twice with the leader's response in any given trial block. Thus the deviation from chance imitation score was calculated by recording the absolute difference between the number of imitative responses made by *S* in a trial block and the value 2. These scores could range from 0 to 2 within a trial block, and total scores could range from 0 to 12. While the total imitative responses measure (*b*) is a meaningful index of the tendency to imitate an adult's behavior, it does not clearly reflect the process of acquiring the imitative habit. The hypothetical *S* who at the outset of training attends to the leader's response as a cue for his own behavior, but begins by making the opposite response and later shifts to the appropriate imitative response, would over the entire training series obtain a total imitation score not differentiable from that of a second hypothetical *S* who disregards the leader's response as a cue and continues for 24 trials to respond at chance level. The deviation from chance imitation score does, however, reflect such differences. The first hypothetical *S* described above should obtain a considerably higher deviation from chance score than the second *S*. It follows then that where the tendency to imitate directly is of focal interest, the total imitative responses score (*b*) is appropriate; on the contrary, where the tendency to attend to the leader's response as a cue for behavior is of focal interest, the total deviation from chance imitation score (*c*) is appropriate. Because these two measures are derived from the same basic data, they are related. The correlation coefficient of only .58 between the two sets of scores indicates, however, that they are somewhat different measures.

An additional possible measure of the rate of acquisition of the imitative habit, the trials to criterion measure, proved useless. The number of trials required to meet the predetermined criterion ranged from 7 for 3 *S*s to infinity (never attained criterion during the training sequence) for 13 *S*s. Because training was discontinued before learning was complete for over 40% of the *S*s, an artificial ceiling was placed on the trials to criterion measure, and it therefore was not regarded in analyzing the data.

Results and Discussion

Table 1 summarizes the behavioral measures for each subgroup in the basic design: (*a*) behavior on Trial #1; (*b*) mean total number of imitative responses; and (*c*) mean total deviation from chance imitation.

Behavior on the Initial Trial

The findings with respect to imitation on the initial trial of task corroborate Miller and Dollard's observation of a preponderant tendency toward nonimitation. Of the 32 *S*s, 9 imitated the leader on the first trial, a percentage proportion of 28.1% as compared with 22.5% in the Miller and Dollard experiment (Miller & Dollard, 1941: Ex-

TABLE 1

Behavioral Measures According to Subgroups in the Basic Experimental Design

Subgroup[a] ($n = 4$)	Measure		
	Number of Imitative Responses on Trial 1	Imitative Responses (Mean)	Deviation from Chance Imitation (Mean)
FYM	0	11.75	6.75
FOM	3	20.25	9.25
FYF	0	11.75	7.25
FOF	1	12.25	8.25
MYM	2	17.50	7.50
MOM	1	13.50	5.50
MYF	1	17.25	6.75
MOF	1	9.75	6.75
Total sample	9	14.25	7.25

[a] See text for code labels for subgroups.

periment V). In spite of the steps taken in this experiment to conceal from the follower information about the leader's reward, the initial predisposition of *S*s to make a nonimitative response in situations of this type nevertheless occurred. A nonparametric analysis of the data revealed no significant differences among the eight subgroups. Furthermore, collapsed combinations of subgroups across the major experimental variables yielded only one significant difference in initial imitative tendency. Older females tended to make significantly more initial imitative responses than younger females, a difference which was significant at the .05 level of confidence by the Fisher exact test.

Total Imitative Responses

The two quantitative experimental measures, total imitative responses (*b*) and total deviation from chance imitation (*c*), were each subjected to a mixed analysis of variance, with sex of *S*, age of *S*, and sex of leader constituting the "between" effects, and trial blocks constituting a "within" effect.

The data summarizing the occurrence of imitative responses yielded total scores ranging from 0 to 24, divisible into six blocks of four trials, numbered for convenience Blocks I through VI. The differences among the six blocks were significant at the .001 level of confidence (see Table 2), indicating that learning of the imitative response occurred over the training series. This then constitutes a replication of the Miller and Dollard experiment. The sharpest increase in amount of imitation occurred between Blocks II (Trials 5-8) and III (Trials 9-12), the t ratio of 2.617 ($p < .02$) indicating the only significant difference between adjacent trial blocks. That the major portion of learning occurred between Blocks I and IV is indicated by the fact that while both the differences between I and IV ($p < .01$) and between Blocks I and VI ($p < .01$) are significant, the differences between Blocks IV and VI is not signifi-

TABLE 2

Summary of the Analysis of Variance of Total Imitative Responses

Variable	*df*	*ms*	*F*
Between *S*s:			
Sex	1	0.33	—[a]
Age	1	0.52	—
Leader Sex	1	12.00	2.00
Sex × Age	1	35.03	5.84*
Sex × Leader Sex	1	1.34	—
Age × Leader Sex	1	11.02	1.84
Sex × Age × Leader Sex	1	1.68	—
Between error	24	6.00	
Within *S*s:			
Trials	5	6.66	7.30**
Sex × Trials	5	0.35	—
Age × Trials	5	0.66	—
Leader Sex × Trials	5	0.59	—
Sex × Age × Trials	5	0.40	—
Sex × Leader Sex × Trials	5	0.52	—
Age × Leader Sex × Trials	5	0.73	—
Sex × Age × Leader Sex × Trials	5	0.40	—
Within error	120	0.91	

* $p < .05$.
** $p < .01$.
[a] *F* ratio less than unity.

cant. These data suggest that if an *S* is likely to learn the imitative habit in this paradigm, he is likely to make the greatest gains in potential prior to the sixteenth training trial.

None of the simple effects (age of *S*, sex of *S*, sex of leader) showed significant effects on total imitative behavior. Although one might expect certain general differences in a tendency toward matched dependent behavior between older and younger *S*s, these data provide no evidence of such differences. For purposes of interest, however, it might be noted that there was a slight (but insignificant—p between .10 and .20) tendency for all *S*s to imitate the male leader more than the female leader. Inspection of the data suggests that this differential tendency is greater among older *S*s than among younger *S*s. Although such differences would be interesting in connection with study of the progressive differentiation of adult sex types and identification, these data warrant no conclusion. Since it is possible that this insignificant trend may be an artifact peculiar to the individual male and female leaders employed in this study, these effects merit further experimental attention.

The interaction between sex of *S* and age of *S* significantly affected the total imitation score ($p < .025$). Within the sample of male *S*s, the mean imitation score for younger *S*s (MY = 17.38) was significantly greater than for older *S*s (MO = 11.63). On the other hand, within the sample of female Ss, this mean for the younger *S*s (FY = 11.75) was significantly smaller than for the older *S*s (FO = 16.25). These differences, significant by t test at beyond the .01 confidence level, would seem consistent with the general assumption that males, in their early training, are progressively taught to adopt a more independent role in behavior, while females are taught to adopt a more reticent, submissive, dependent role. Among the older *S*s, the females imitated significantly more often than the males ($p < .01$); however, it is surprising that these sex differences were significant and in the opposite direction for younger *S*s. These observations involve the comparison of subgroups containing eight *S*s, and although the trend is replicated in the same fashion with both male and female leaders, additional cross-validation is warranted.

The interaction effects of the "between" variables (sex of *S*, age of *S*, sex of leader) with trial blocks indicate the extent of difference among subgroups in the forms of learning curves for acquisition of the imitative habit. None of these interactions were statistically significant.

Deviation from Chance Imitation

The deviation from chance imitation score is an index not of the tendency to imitate adults, but of the extent to which the adult leader's response appeared to function as a determinant of *S*'s behavior.

The within effect, trial blocks, fell short of the required .05 significance level. Thus, while the data indicate that the imitative response per se can be taught experimentally, the disposition to attend to the

TABLE 3

SUMMARY OF THE ANALYSIS OF VARIANCE OF TOTAL DEVIATION FROM CHANCE IMITATION SCORES

Variable	*df*	*ms*	*F*
Between *S*s:			
Sex	1	2.08	1.39
Age	1	0.19	—[a]
Leader Sex	1	0.00	1.68
Sex × Age	1	2.52	—
Sex × Leader Sex	1	0.08	—
Age × Leader Sex	1	0.02	—
Sex × Age × Leader Sex	1	1.02	—
Between error	24	1.50	
Within *S*s:			
Trials	5	0.97	2.14
Sex × Trials	5	0.17	—
Age × Trials	5	0.30	—
Leader Sex × Trials	5	1.01	2.23*
Sex × Age × Trials	5	0.48	1.07
Sex × Leader Sex × Trials	5	0.10	—
Age × Leader Sex × Trials	5	0.36	—
Sex × Age × Leader Sex × Trials	5	0.26	—
Within error	120	0.45	

* $p < .05$.
[a] F ratio less than unity.

leader's response as a cue for behavior is not effectively learned in such experimental training. The interaction between leader's sex and trial blocks was significant at the .05 level of confidence. This significance is due altogether to that group of *S*s trained to imitate a male leader. While none of the differences between trial blocks is significant for *S*s trained to imitate a female, the mean differences between Blocks I and IV (0.81 versus 1.25), between Blocks I and V (0.81 versus 1.63), and between Blocks I and VI (0.81 versus 1.56) are significant at or beyond the .02 confidence level by *t* test for *S*s trained to imitate male leaders. These findings are interesting when the fact is considered that the most frequent adult "teachers" in the learning experience of young children are females: the mother, the preschool teacher, etc. The deviation from chance imitation score on Block I only for imitation of a female leader (1.38) is significantly greater than for imitation of a male leader (0.81) at the .01 level. This suggests that the disposition to attend to the adult female's response as a social cue is considerably greater at the outset of training. Thus, although the total scores over the entire training series do not differ for *S*s trained to imitate males and *S*s trained to imitate females, the data indicate that attention to the female's response is likely to have been learned prior to the experimental training, whereas attention to the male's response is fairly effectively learned during the training sequence.

None of the simple between-subjects effects reached significance.

IMITATIVE BEHAVIOR AND INTELLIGENCE

The component variables defining intelligence are ill-specified, but in any learning situation there is some likelihood that intellectual factors operate to some extent in determining an individual's performance. For this reason it is imperative that before other variables are explored as potential determinants of imitative behavior, the role of intelligence should be clarified.

It is conceivable that the highly intelligent *S* might in this situation be alert to cues of all categories (both social and environmental), thus shrewdly "catching on" to the fact that imitative behavior is reinforced and consequently deriving high scores both in terms of total imitative behavior and in terms of deviation from chance imitation. Such an analysis yields an hypothesis of direct relationship between observed imitation and intelligence. On the other hand, it is equally feasible that the highly intelligent *S* might be more likely to attempt to use the leader's behavior as a cue for social learning, directing attention to the environmental cues, and pursuing false leads in attempts to utilize the leader's response as an instrument to solution of the problem on the basis of environmental cues. Such an *S* should suffer greater interference from the randomized color and position cues, leading to slower learning and lower scores both in terms of total imitative behavior and deviation from chance imitation. This analysis yields an hypothesis of inverse relationship between imitation and intelligence.

Method

For 26 of the original sample of 32 *S*s, Binet intelligence test scores were available. During the course of the academic year, graduate students in a course dealing with the psychological appraisal of young children customarily utilize children in the laboratory preschools as practicum *S*s for administration of Form L of the Stanford-Binet tests of intelligence (Terman & Merrill, 1937). When such administrations are found by the instructor to meet requisite standards, the scores are recorded on the permanent file of each preschool child. The scores used in this exploration were derived from this source. The recency of the test scores varies from *S* to *S*, but in no case were scores older than 18 mo.

The mean IQ for the sample of 26 children was 120.9, with scores ranging from 87 to 150. The scores were split at the median (119), providing for a High Intelligence group and a Low Intelligence group. In order to control for the effects of the basic experimental variables observed to affect imitation (sex and age), both subsamples included four younger females, three older females, four younger males, and two older males. There were no systematic relationships between intelligence and age or sex.

The imitative response scores were submitted to a simple mixed analysis of variance with IQ level constituting the between-subjects component and trial blocks constituting the within-subjects component. Since deviation from chance scores did not vary over trial blocks, the difference between these scores for the two IQ levels was tested by simple *t* test.

Results and Discussion

The analysis of variance of imitation scores indicated no relationship between number of imitative responses and intelligence level. Likewise, the *t* test indicated no difference in total deviation from chance scores for the two intelligence levels. The product-moment correlation between Binet IQ scores and total deviation from chance scores was −.11, which is not significantly greater than zero. However, the correlation between total imitation scores and Binet IQ scores was +.42, which is significant at the .05 level of confidence.

These findings suggest that predispositional attention to adult behavior is unrelated to tested intelligence, but that direct imitation of the adult's behavior bears a slight relationship to intelligence. Such a relationship might be expected as a general function of learning ability rather than of tendency to imitate adults. Since the deviation from chance imitation measure is apparently free of intelligence effects, and since total imitation measure, which is more directly a function of learning ability, is only moderately correlated with intelligence, it is unlikely that intellectual factors alone can account for a considerable amount of the observed variance among individuals in imitative behavior.

CHILD REARING ANTECEDENTS OF IMITATIVE BEHAVIOR

This phase of the experimental study of imitative behavior was directed toward the exploration and analysis of individual differences in behavior under controlled laboratory circumstances as a function of child rearing antecedents. The position adopted in executing these studies assumes that individual differences in the predisposition to behave in an imitative or matched dependent manner arise at least in part from differential patterns of experienced reinforcement in related situations. Inasmuch as the present experimental design assesses the extent to which a preschool child imitates an adult, one of the most consequential related areas of experience should be that of parent–child interaction, the prototypical situation in which the child learns most of his habits of interaction with adults. It would be expected that the child reared under circumstances in which his blind imitation of adult behavior is consistently rewarded and his devious attempts at independent exploration of the consequences of his own independent behavior punished should be predisposed to a general tendency to follow the behavior of adult figures.

Research findings in this area are largely consistent with such an hypothesis, although the methods of assessment of variables in such research are often imprecise and the variables themselves ill-defined. While certain studies have made use of direct assessment of parental attitudes and practices (Champney, 1941; Radke, 1946; Sears, Whiting, Nowlis, & Sears, 1953; Shoben, 1949), most have utilized case histories (Newell, 1936) and judgmental ratings by the researcher (Baldwin et al., 1945; Knight, 1933; Lasko, 1952) for assessment of the parental variables under investigation. Similarly, while a few studies have utilized controlled observation (Hattwick, 1936; Radke, 1946; Sears et al., 1953), several have used projective assessment (Radke, 1946) and after-the-fact judgmental ratings by observers (Knight, 1933; Lasko, 1952) as techniques for collecting data on child behavior. Radke (1946), in integrating research evidence in this area, has summarized a number of relationships between parental practice variables and child behavior either directly or indirectly relevant to imitation. This summary indicates some consistency among the findings, but also a certain amount of apparent incongruity. Behavior labeled "submissive" has been shown to be related to parental "overprotection" (Grant, 1937; Hattwick, 1936; Newell, 1936) as well as to parental "rejection" (Newell, 1936; Radke, 1946) and "domination" (Radke, 1946; Symonds, 1939). Behavior labeled "noncompliant" has likewise been demonstrated to relate to "rejection" (Meyers, 1944) and parental "autocracy" (Radke, 1946), as well as to parental "submissiveness" (Meyers, 1944). In addition, behavior defined as "independent" and "self-reliant" has been shown to be related to parental "submissiveness" (Symonds, 1939), to a "logical-scientific approach to child rearing" (Grant, 1937), and to "warm interaction with the child" (Hattwick, 1936). While parental "overprotection" and "rejection" may appear on the surface to be incompatible phenomena, Levy (1943) has proposed a distinction between phenotypic rejective *behavior* toward the child and genotypic rejective *feelings* which might lead to compensatory overprotection in overt behavior. Thus, some reconciliation of apparent contradiction among the findings is effected. Nevertheless, without exhaustive analysis of the concrete definition of such variables as "submissiveness," "rejection," and "overprotectiveness" in parents, such research evidence is of doubtful value either to the researcher or to one dealing directly with children.

Hattwick (1936), in a study somewhat comparable to the present investigation, utilized planned and controlled observation of children and of their parents (in home visits) and found that help-seeking behavior in the child is directly related to overprotective practices in the home.

While the present study did not propose the test of explicitly stated hypotheses, it was designed to explore the extent to which previously derived laboratory measures of a predisposition to imitative behavior among preschool children are related to certain variables of parental attitude through the Parental Attitudes Research Instrument (Schaefer & Bell, 1958) and to certain variables of parental practice, through a modification of the Sears, Maccoby, and Levin (1957) maternal interview (Gold, 1958).

Method

Sources of Data

Parental Attitudes. The Parental Attitudes Research Instrument (PARI) is an uncopyrighted research instrument devised by Schaefer and Bell (1958) to assess a number of attitudinal areas focused upon the rearing of children. The instrument itself has been through several modifications by its authors and has been submitted to at least two factor analyses (Schaefer & Bell, 1957; Zuckerman, Ribback, Monashkin, & Norton, 1958). The attitude data utilized here are quantitatively unrefined, in that two slight modifications of Form IV of the PARI were employed in different test contexts. It was nevertheless assumed that the crude measures thus obtained should yield fairly comparable and meaningful scales. The PARI, in one of two forms, was administered to the mothers of 31 of the original 32 experimental *S*s, and to the fathers of 29 of these *S*s. For two mothers and eight fathers, PARI Form IV was administered alone in the spring of 1957; for the remaining 29 mothers and 21 fathers, PARI Form IV was administered in early 1958 within an experimental test battery devised by Chantiny, Lovell, and McCandless (1956). This battery, in addition to a very slightly reworded modification of PARI Form IV (which made it appropriate for either parent by the substitution of the word *parent* for the words *mother* or *father* wherever they occurred), also contained adaptations of the California F Scale and the Taylor Manifest Anxiety Scale. Each item in the PARI presents a statement of an attitudinal position, and requests the parent to indicate the extent of his agreement in one of four categories: strongly agree (scored 4), mildly agree (scored 3), mildly disagree (scored 2), and strongly disagree (scored 1). Thus, on each subscale of five items, scores ranging from 5 to 20 were recorded. The subscales explored in this research, selected from the total battery included in the PARI, were:

1. Comradeship and Sharing
2. Autonomy of the Child
3. Suppression of Sexuality
4. Avoidance of Communication
5. Encouraging Verbalization
6. Breaking the Will
7. Equalitarianism
8. Deification of the Parent
9. Fostering Dependency
10. Intrusiveness
11. Strictness
12. Suppression of Aggression

The labels given these scales by the authors of the PARI are generally descriptive of their content. The Suppression of Sexuality and Suppression of Aggression scales index permissiveness in the areas of sexual and aggressive behavior. Avoidance of Communication and Encouraging Verbalization are scales which measure the extent to which the parent believes the child's thoughts and opinions should be expressed. The Intrusiveness scale assesses the parent's intrusion upon the child's privacy of thought and activity. The first five scales listed above loaded most heavily on Factor I, Suppression and Interpersonal Distance, one of five emergent factors in a factor analysis of an early form of the PARI (Schaefer & Bell, 1957). Scores on these scales were summed algebraically, in terms of their direction of loading, to yield a combined Factor I score for each *S*.

Maternal Practices. A second set of data involving maternal child rearing practices was explored similarly. In connection with other ongoing research at the Iowa Child Welfare Research Station, a modification of the Sears et al. (1957) direct maternal interview had been undertaken (Gold, 1958). These data were available for 22 of the original 32 experimental *S*s. The procedure consisted of a structured interview of approximately 90-min. duration with the mother, executed by two female interviewers by prior appointment at one of the preschool buildings, and recorded with the mother's knowledge for later transcription. The interviewers underwent a preliminary period of training to maximize comparability in interviewing technique, and the transcribed interview was rated on predefined scales by two graduate students. The material included in the interview was adapted directly from the Sears interview, and the rating scales employed were identical. In rating the material, the ratio of rating agreements to total ratings for the two judges was .91. Discrepancies of two or more rating points were discussed to agreement, while discrepancies of one point were simply averaged. All ratings were converted by linear transformation to a standard nine-point scale, and were used in that form in the analysis of these data. The following 19 subscales (identified by the number assigned them by Sears et al.) were studied:

II-23	Amount of affection interaction with the child
II-43	Demands for table manners
II-45	Praise for good behavior at the table (high score-none)
II-53	Modesty permissiveness
II-56	Masturbation permissiveness
II-58	Sex-play permissiveness
II-61	Standards for neatness-orderliness
II-62	House and furniture restrictions
II-66	Noise restrictions
II-73	Standards for obedience
III-11	Disapproval of dependency
III-13	Affectionate demonstrativeness toward the child
III-14	Time to play with the child (high score-little time)
III-23	Praise for playing nicely (high score-none)
III-30	Permissiveness for aggression toward parents
III-31	Punishment for aggression toward parents
III-36	Use of tangible rewards
III-47	Extent of use of deprivation of privileges
III-57	Extent of use of reasoning

Milton's (1958) factor analysis of the Sears interview material yielded three factors which appear to be descriptive of parent-child interaction patterns. In order to derive a factorial score for each factor, the scores on individual scales which loaded ±.35 or greater were combined algebraically in terms of their direction of loading. Using this procedure, Factor S (Strictness–permissiveness) includes Scales II-43, II-53, II-56, II-58, II-61, II-62, II-66, III-11, III-30, and III-31. Factor W (Warmth of the mother–child relationship) includes Scales II-23, II-45, III-11, III-13, III-14, and III-57. Factor R (Responsible child rearing orientation) includes Scales II-45, II-61, II-73, III-23, III-36, and III-47.

Child Behavior. The measures of child behavior utilized here are derived directly from the laboratory study of imitative behavior. Three measures were analyzed: (*a*) total imitative responses (number of imitative responses in the 24-trial training series); (*b*) total deviation from chance imitation (absolute difference between the total number of imitative responses made and the number expected by chance); and (*c*) initial tendency to imitate (behavior on Trial #1 of the training series). For the first two measures, product-moment coefficients were calculated as indices of the extent of correlation between child behavior and parental attitudes and practices. Although a number of factors lead to the expectation of nonlinear relationships in an analysis of this type, the linear product-moment coefficient is apparently a serviceable index for these data. The statistic epsilon, a derivative of the correlation ratio (eta) which is unbiased by sample size and coarseness of grouping (Kelley, 1935) was calculated for each of the relationships explored, and in no case was the hypothesis of linearity of regression negated. Hence the correlation indices reported here are such that if the coefficient is statistically significant, the relationship between the two variables concerned may be assumed to be roughly linear.

In considering the third variable of child behavior, initial tendency to imitate, point-biserial correlation coefficients were calculated, since the criterion variable (imitative first response versus nonimitative first response) is a true dichotomy.

Results and Discussion

The results of this exploration are not easy to integrate. The predictor variables employed (PARI subscales and interview rating scales) are highly variable measures of broadly defined variables. Furthermore, the criterion variables studied (laboratory behavioral scores) are measures of relatively specific behaviors of the child. The samples are small, and a relatively large coefficient of correlation is required for the attainment of statistical significance. In view of the large number of correlation coefficients calculated, a certain number of significant relationships are probable, allowing possible Type I statistical errors. Since the study rests upon no explicit predictive hypotheses, only those relationships which proved to be statistically significant are discussed here. The complete table of correlational relationships explored is included as a supplementary appendix. Tables 4 and 5 summarize the empirical relationships which attained statistical significance, and which constitute the basic data for interpretation.

Total Imitative Responses

One of the criterion measures derived from the laboratory experiment is the total number of imitative responses made (or projected) during the 24-trial training sequence. This measure upon analysis seems to be the derivative of (*a*) a consistent tendency to emulate the behavior of adults, (*b*) attention to the adult's response as a determining cue for the child's own behavior, and/or (*c*) efficient integration of experienced reinforcement in development of the so-called "imitative habit." That is, there are at least three components affecting the magnitude of this score.

Parental Attitude Measures. For the female *S*s in this study, the total imitative responses measure was related to one maternal attitude measure and two paternal attitude measures. The maternal Suppression of Aggression scale was correlated +.52 ($p < .05$) with this measure. While the content of these items on the test instrument is directed toward restriction and limitation in a specific area of behavior, the scale probably reflects a somewhat more generalized attitude of strict nonpermissiveness in dealing with children. The paternal administration of the PARI yielded two scales which correlated significantly with total imitation: Avoidance of Communication ($r = +.53$; $p < .05$) and Autonomy of the Child ($r = -.55$; $p < .05$). The former scale contains items which generally suggest rejective denial of the child's initiation of verbal interaction, and the latter is an index of the extent to which independent exploration by the child is approved and fostered.

The prediction of laboratory imitation of boys from parental attitude measures is more ambiguous than for girls. Only one parental attitude scale was significantly related to behavior for the male *S*s: the paternal Strictness scale, which correlated +.57 ($p < .05$) with total imitative responses.

For the total sample (combined males and females), only one parental attitude scale was related to total imitation: the paternal Autonomy of the Child scale correlated −.37 ($p < .05$) with this measure, but the relationship is accounted for largely by the strength of the correlation among females. The correlation between these two measures for males was only −.14, which is not significantly different from zero.

There are undoubtedly meaningful differences in predictive validity between maternal and paternal child rearing attitudes. The correlations between attitudinal measures for pairs of parents was uniformly low within the present sample. Although one might expect useful information about the identification process to emerge from comparative analysis of the behavior of male and female children with respect to the child rearing attitudes of their mothers and fathers, the present study afforded little opportunity for such exploration.

Maternal Practices Measures. The observed relationships between the child's laboratory behavior and child rearing practices reported after-the-fact by the mother were generally consistent with the emergent pattern of relationships between behavior and parental attitudes. The schematic design of the Sears type interview is one which affords relatively more precision in definition of the measured variable than does the PARI, but it is at the same time more open to difficulties of distortion. The interview situation is one in which distortion for interpersonal approval is maximal; furthermore, the intervening process of judgmental rating affords the introduction of "personal

TABLE 4

CORRELATIONS BETWEEN PARI SCALES AND LABORATORY IMITATION (Pearson r)

PARI Scale	Behavioral Measure					
	Total Imitation			Deviation from Chance Imitation		
	Boys	Girls	Total	Boys	Girls	Total
Maternal PARI						
Suppression of Sexuality	−.19	+.32	+.08	−.13	+.58*	+.17
Suppression of Aggression	−.30	+.52*	+.03	.00	+.57*	+.10
Breaking the Will	−.30	−.21	−.21	−.61*	−.30	−.50
Factor I: Suppression and Interpersonal Distance	+.34	−.04	+.10	+.41	−.51	+.04
Paternal PARI						
Autonomy of the Child	−.14	−.55*	−.37	+.28	−.60*	−.12
Avoidance of Communication	−.08	+.53*	+.14	−.18	+.46	−.01
Intrusiveness	+.28	+.10	+.16	+.79**	+.18	+.30
Strictness	+.57*	−.12	+.15	−.19	+.65*	+.13
Factor I: Suppression and Interpersonal Distance	+.07	−.50	−.24	+.38	−.45	−.01

* $p < .05$. ** $p < .01$.

TABLE 5

CORRELATIONS BETWEEN MATERNAL INTERVIEW SCALES AND LABORATORY IMITATION (Pearson r)

Interview Scale	Behavioral Measure					
	Total Imitation			Deviation from Chance Imitation		
	Boys	Girls	Total	Boys	Girls	Total
Permissiveness for aggression toward parents	+.12	−.81**	−.57**	+.27	−.40	−.11
Punishment for aggression toward parents	−.15	+.68**	+.35	−.60	+.24	−.11
Amount of affectionate interaction with child	.00	+.14	+.06	−.37	+.56*	+.21
Factor S: Strictness–permissiveness	−.41	+.75**	+.25	−.53	+.65*	+.04
Factor W: Warmth of mother–child relationship	−.05	−.29	−.22	−.14	−.52	−.42
Factor R: Responsible child rearing orientation	−.21	−.06	−.12	−.32	−.04	−.21

* $p < .05$. ** $p < .01$.

equation" errors. Nevertheless, the findings emergent in this exploration appear to be relatively consistent across both instruments.

None of the interview rating scales was significantly related to the laboratory behavior of the male *S*s. This fact appears to lend further credence to the assumption that the behavior of young males in situations of this type may be somewhat less amenable to simple analysis than the behavior of young females.

For the subsample of 13 female *S*s for whom maternal interview data were available, three individual rating scales and one of the combined factorial scales were correlated with the total number of imitative responses in the laboratory. This score correlated −.81 ($p < .01$) with Scale III-30 (Permissiveness for aggression toward parents) and +.68 ($p < .01$) with Scale III-31 (Punishment for aggression toward parents). These two scales were mutually correlated in Sears' initial sample and are evidently interrelated aspects of parent–child interaction (Sears et al., 1957). While they tap an area concerned with the maintenance of parental authority over the child, they are in all likelihood specific aspects of a more general tendency for the parents to maintain relatively strict control of the child. This interpretation is suggested by the strong loadings of these two scales (−.51 and +.43) on Factor S (Permissive–strictness) in the Milton factor analysis of the Sears data (Milton, 1958). In line with this interpretation, the Factor S combined score is also correlated with total imitation ($r = +.75$; $p < .01$).

Since the correlations between behavior and interview data for boys were uniformly low, in only one case did the correlation between total imitation and maternal practices attain significance for the combined sample (males and females). Scale III-30 (Permissiveness for aggression toward parents) was correlated −.57 ($p < .01$) with this measure, but the magnitude of this relationship is due to the correlation within the female subsample. The same correlation for males only was +.12.

Deviation from Chance Imitation

A second, perhaps less ambiguous, behavioral measure from the laboratory experiment concerned the extent to which the child's behavior appeared to have been determined by the adult leader's response. This score indicates the absolute difference between the number of imitative responses made by the child and the number expected by chance during training. It may be construed as an index of dispositional cue selection; that is, of a general tendency to attend to "social cues" when pitted against direct environmental sensory stimuli as potential determinants of behavior. The fact that this score did not increase significantly over the training series implies that reinforced training had little effect in altering such a dispositional tendency.

Parental Attitude Measures. For the females, this social attention measure was significantly related to two maternal and two paternal attitude scales. Correlations of −.57 with the maternal Suppression of Aggression scale and +.58 with the maternal Suppression of Sexuality scale were both significant at the .05 confidence level. While both of these attitudinal scales focus upon specific behavioral areas, they are probably homogeneous in reflecting generalized nonpermissiveness in child rearing. The deviation from chance imitation measure also correlated +.65 ($p < .01$) with the paternal Strictness scale, and −.60 ($p < .05$) with the paternal Autonomy of the Child scale.

For the male *S*s, two parental attitude scales were correlated with deviation from chance imitation. This measure correlated +.79 ($p < .01$) with the paternal Intrusiveness scale, which contains items describing severe limitation of the child's privacy in thought and behavior. In addition, deviation from chance imitation correlated −.61 ($p < .05$) with the maternal Breaking the Will scale. This scale contains items which describe harsh and punitive control of the child, especially through fear of the parent. While the body of evidence yielded in this exploration suggests that parental strictness may foster dependent attentiveness to adults for behavioral cues, it may be possible that high scores on the Breaking the Will scale indicate a parental position leading to a generalized distrust of adults by the child. On the surface, this observed relationship appears inconsistent with the remainder of the emergent body of evidence described here; however, this analysis of the attitudinal position described in this particular scale would generate a prediction of independent behavior consistent with the observed relationship as well as with the remainder of the present findings. The corresponding correlation within the female subsample was $-.30$ (nonsignificant), and the combined sex correlation between deviation from chance imitation and the maternal Breaking the Will scale was $-.50$ ($p < .01$).

Maternal Practices Measures. The extent to which the child's behavior deviated from chance-expected imitation in the laboratory correlated $+.56$ ($p < .05$) with Interview Scale II-23 (Amount of affectionate interaction with the child) for girls. This scale represents the extent to which the mother reported open affectionate display with her child, and it may possibly be related to general "babying" of the child and attempts to maintain his dependent relationship with his parents. The deviation from chance imitation scores did not correlate significantly with either of the two scales concerned with aggression toward parents (as did the total imitation score), but correlated $+.65$ ($p < .05$) with Factor S (Strictness–permissiveness).

TABLE 6

CORRELATIONS BETWEEN INITIAL TRIAL BEHAVIOR AND MATERNAL INTERVIEW SCALES
(Point-biserial *r*)

Interview Scale	Correlation with Imitation on Initial Trial
Disapproval of dependency	+.43*
Factor S: Strictness–permissiveness	+.14
Factor W: Warmth of mother–child relationship	−.52*
Factor R: Responsible child rearing orientation	−.42

* $p < .05$.

Initial Tendency to Imitate Adult Behavior

While the experimental situation was designed for the training of imitative behavior by consistent reinforcement, the data already discussed provide some evidence that there occur meaningful individual differences in the course of acquiring the learned imitative habit. Aside from training effects, however, the child's initial behavior on the first trial of the experimental series may be construed as a one-shot measure of his initial predisposition to imitate the behavior of adults. Of the 32 experimental *S*s, 9 (4 females and 5 males) imitated the response of the adult leader on the first trial. Although this is an ambiguous observation, in that it may be chance determined, it is the only measure in this experiment which is completely free of experimental reinforcement effects. The *S*s were split into dichotomous groups on the basis of whether they made the same (imitative) response or the opposite (nonimitative) responses to that of the adult leader on the first choice of the experimental series, and point-biserial correlation coefficients with each of the parental attitude and practice measures were calculated. None of the correlations with parental attitude measures was significant, but this predisposition was correlated $-.52$ ($p < .02$) with Factor W (Warmth of the mother–child relationship), and $+.43$ ($p < .05$) with Scale III-11 (Disapproval of dependency) on the maternal interview. While the Factor W score may be characterized as a measure of a warm, democratic, affectionate but not possessive, relationship between the mother and the child, Scale III-11 is perhaps the most direct available measure of the consistency of reinforcement for dependent behavior. Miller and Dollard (1941, p. 121) suggest that the initial tendency to imitate may be largely due to "environmental conditions which have rewarded the (organism) for learning to imitate under conditions similar enough to the experimental set-up so that generalization can occur." One would expect, then, that consistent parental reinforcement of dependent behavior should be directly related to imitation on the first trial of this experimental training series. These data, however, indicate the reverse. High scores on Scale III-11 reflect fairly consistent maternal rejection of and punishment for the child's dependent overtures. It is possible that "dependency" as implicitly defined within the context of the interview does not afford a realistic measure of maternal reinforcement for actual matched-dependent behavior; furthermore, it is possible that the present experimental situation may not afford ready generalization from the conditions of maternal reinforcement of dependent behavior. However, the fact that the correlational relationship observed here is significant and in the opposite direction suggests a need for reconsideration and alternative analysis of the effects of consistent parental rejection of the child's dependency. It may be that Scale III-11 measures a specific manifestation of general parental punitiveness, in which case the measured variable may be closely related to other measures of nonpermissiveness in such areas as aggression and sexual behavior. On the other hand, it is possible, as Sears has suggested (Sears et al., 1953, 1957), that conflict in the area of dependency is a major issue in the social development of children. Consistent parental rejection of and punishment for dependent behavior may lead to a false and unrealistic kind of independence on the child's part. Such "forced independence" may lack elements of confidence and self-assurance that are vital to mature independence. It is interesting to note that scores on Factor R (Responsible child rearing orientation), which Sears has tentatively described as a measure of "earnest and responsible" orientation to child rearing duties, the mother's "regard for herself," and devotion of time and effort to administration of "appropriate rewards and punishment in the training process" (Sears et al., 1957, p. 475), are correlated $-.42$ ($p = .08$) with the child's initial tendency to imitate. This coefficient is not statistically significant, but its magnitude implies that a reconsideration of such variables in connection with dependency conflict might provide fruitful information.

Imitative Behavior and the Sibling Constellation

Research concerned with the development of personality and social behavior has rather consistently provided evidence of clear-cut behavioral differences related to the individual's ordinal position among his siblings. Although the methods employed in such studies vary widely, they most frequently make use of behavioral ratings either by the experimenter (Dean, 1947; Goodenough & Leahy, 1927; Lasko, 1954) or by the mother or a teacher dealing directly with the subject population (Kawin, 1934; Koch, 1955). The descriptive pattern which emerges to characterize the oldest child in a family of several siblings appears most consistent from study to study. The oldest child has been described as less aggressive (Cobb, 1943; Goodenough & Leahy, 1927; Kawin, 1934), less self-confident (Cobb, 1943; Goodenough & Leahy, 1927; Koch, 1955), reluctant to make decisions (Dean, 1947), lacking in leadership qualities (Cobb, 1943; Goodenough & Leahy, 1927; Roberts, 1938), and suggestible and gullible (Goodenough & Leahy, 1927). In interpreting these findings, some significance has been attached to the fact that the oldest child occupies the peculiar position of having to give up a once-held "only child" status. Although speculative interpretation has ranged from Adlerian assumptions about the increased press toward aggressive striving in younger siblings (Adler, 1930) to analyses of the different social learning situations encountered by particular ordinal positions (Mussen & Conger, 1956), such discussions are in general concerned most directly with relationships among siblings rather than with differential parent–child relationships which occur as a function of ordinal position. In considering the marked inconsistency of research findings concerning contrasts between the only child and the child with siblings, Mussen and Conger (1956, p. 343) conclude that " 'onliness' itself is not related to any specific type of personality structure or problem behavior. The child's familial background and his relationships with his parents probably influence his personality development much more than the presence or absence of siblings."

The findings described earlier concerning relationships between imitative behavior in the laboratory and parental attitudes and child rearing practices suggest that the parent–child relationship may play a critical role in determining the manner in which a child learns to make use of his observations of the behavior of other people in his own social learning. There was evidence that restrictive, nonpermissive, and autocratic, as well as extremely close, intrusive, and overprotective parental behavior patterns were related to a tendency for the child to utilize social cues as blind directives for behavior rather than as instruments for social learning. It was concluded that such "interventive" practices on the part of parents may in fact tend to restrict the child's independent exploration of his environment and thus increase his reliance upon others for direction. From this, one might speculate that the likelihood for such restrictive interference to occur in a child's early experience may be considerably greater if he is the first-born or only child, but diminished if he is one of several children among whom parental attention must be divided. In addition to the simple realistic division of attention, there may be other factors which lessen the amount of overprotective or intrusive restriction directed by parents toward subsequent children after the first. The birth of the first child into a family is a novel experience for both parents, and usually a much hoped for event. Sears, Maccoby, and Levin (1957) cite evidence that the mother's reaction to pregnancy tends to be significantly more positive with the first child than with later children. Such lessened enthusiasm and eagerness, as well as increased demands on the mother with more than one child, may differentiate the relationship of a first child to his parents from that of subsequent children. If such assumptions as these are true, they may provide a link of information contributing to the explanation of certain observed behavioral differences related to ordinal position in the family.

While research studies in this area customarily differentiate *only* from *oldest* children (Guilford & Worcester, 1930), the two groups are considered in combination here. Although the pattern of sibling interaction is categorically different in the two cases, there are critical similarities with respect to the particular variables of parent–child interaction with which the present study is directly concerned.

By extending this analysis of parent–child relationships, one might further hypothesize that the greater the chronological separation between a child and his next older sibling, the more that child will resemble an oldest child in terms of the relationship pattern under consideration. A child born into a family at a time when his next older sibling is still in diapers, not yet feeding himself, perhaps not yet toilet trained, and still in many ways requiring a considerable amount of direct attention from the parents, might be expected to receive a relatively smaller amount of protective or restrictive attention from his parents than will the child separated from his next older sibling by a wider span of years. The pattern of parent–child relationships which is assumed to be critical here is likely to be of greatest consequence if practiced at a time when the child's first attempts at independent exploration occur. Such attempts probably begin to be significantly numerous with increased efficiency in locomotion at the end of the first year and early in the second year. They are, of course, likely to continue over a number of years.

Method

*S*s for this study were the 32 preschool children employed in the basic experimental training of imitative behavior. From permanent biographical records, information concerning the sibling constellation in the home of each *S* was obtained, and the sample was divided into three groups: (I) oldest and only children ($n = 9$), (II) youngest children ($n = 13$), and (III) middle children or children with both older and younger siblings ($n = 10$). In view of the previously observed effects of age and sex in interaction upon imitative behavior, a balanced distribution within each group with respect to age and sex was achieved, yielding three groups of eight *S*s each.

The total imitative response scores were subjected to a mixed analysis of variance, with sex of *S* and ordinal position as between-subjects factors and trial blocks as a within-subjects factor. Similarly, the total deviation from chance imitation scores were submitted to a factorial analysis of variance with respect to sex and ordinal position.

The second hypothesis, concerning expected behavioral differences between closely and widely spaced younger siblings, was explored by comparing the total imitation and total deviation from chance scores for these two groups after preliminary analysis of variance had indicated no differential effects over trial blocks. Within the two groups of children having older siblings (II and III), *S*s were reclassified into two groups on the basis of chronological spacing or the age interval between *S* and his next older sibling. An interval of two and one half years was arbitrarily selected as the cutting point. Children born sooner than 30 mo. after their next older sibling were classified as closely spaced (10 *S*s), and children born later than 30 mo. after their next older sibling were classified as widely spaced (13 *S*s). Again, in order to achieve proportional distribution of sexes within the two groups, three *S*s were eliminated at random from the widely spaced group, resulting in two groups of 10 *S*s (4 females and 6 males) each.

Results and Discussion

The analysis of variance of the imitative response data indicated no significant effects of either sex or ordinal position (either independently or in interaction with trial blocks) upon learning the imitative response in the laboratory. The mean total imitation score for oldest and only children (17.00) is considerably higher than the similar scores for middle (11.25) and youngest (12.00) children, but these differences are not statistically significant.

The factorial analysis of variance of total deviation from chance scores indicated no significant effects due to sex of the child, but the effect of ordinal position was significant at the .05 confidence level. Simple t tests of the differences between the means for the three groups revealed the greatest magnitude of difference between the mean of 8.75 for the oldest and only group and the mean of 5.25 for the middle group ($p < .05$). The mean of 6.75 for the group of youngest children falls between the means for the other two groups, and is not significantly different from either. In magnitude, however, it is nearer the mean for the middle group than to that for the first-born group.

Comparison of behavioral scores for widely spaced and closely spaced children with older siblings revealed no significant differences. However, the total deviation from chance imitation scores for these two groups yield some indirect support for the original hypothesis when compared with scores for the first-born group. The mean deviation from chance score for closely spaced children was 5.50, while that for widely spaced children was 7.60. These means, when compared with the corresponding mean score of 8.75 for first-born children, indicate that widely spaced younger siblings are more like first-born children than they are like closely spaced younger siblings. Although the difference between mean scores for closely and widely spaced children is not statistically significant, the difference between mean scores for the first-born children and the closely spaced younger siblings ($8.75 - 5.50 = 3.25$) is significant by t test at the .02 confidence level. The mean scores for widely spaced younger children and for first-born children

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE OF DEVIATION FROM CHANCE IMITATION SCORES BY SEX AND ORDINAL POSITION

Variable	*df*	*ms*	*F*
Sex	1	13.50	2.29
Ordinal Position	2	24.50	4.16*
Sex × Position	2	10.50	1.78
Within cells error	18	5.89	
Total	23	189.83	

* $p < .05$.

are, in accord with the original hypothesis, not significantly different.

This analysis of behavioral differences in terms of ordinal position in the sibling constellation is based on an assumption that the quality and nature of the parent–child relationship contributes an intervening variable accounting for the observed differences. In order to document such an analysis, one should expect that parents report different practices in interaction with first-born children as compared with subsequent children. Sears et al. (1957, p. 409) have described several differences in maternal practices as a function of the ordinal position of the child under consideration. Restrictions on physical mobility were slightly greater for first-born children, and progressively less for subsequent children. Furthermore, the amount of attention devoted to keeping track of the child was less for middle children than for only, oldest, or youngest children.

With respect to the central variables under consideration here, parental strictness and intrusive overprotection, there are two points of caution to be noted. The division of attention which necessarily accompanies the birth of a second child into a family affects not only the second child but the first as well. It may be true that the second child is unlikely to experience at any time the devotedly protective attention experienced by the only or oldest child. However, it is equally true that the birth of the younger sibling carries with it, in addition to the traumatic loss of this attention, the acquisition of new freedom from such attention, for whatever it might be worth. It is important to consider, then, the relative significance of such autonomous freedom at various stages in the progression of social development. While early learning is in all probability crucial in setting indelible patterns of experienced reinforcement, there may be certain basic skills and responses which are acquired in later development, possibly after the arrival of younger siblings. These later responses may carry equal if not greater import with respect to the eventual achievement of autonomy and independence. Secondly, the intrinsic effects of transition from being the sole object of parental attention to a position of sharing attention and acquiring autonomy may represent a critical sudden change of reinforcement schedule bearing meaningful implications for the first child's social development.

The Generality of Imitative Behavior

The generality of behavioral predispositions in experimental literature is more often implied than stated explicitly. However, in many cases stated principles derived from laboratory observations are extended to encompass more complex forms of behavior without explicit exploration of the generality of the tendency observed in the laboratory. It may be well to recommend a cautious approach to the use of highly sterilized sets of observations in the psychological laboratory as the starting point for speculation about more complex behavioral phenomena.

Miller and Dollard (1941) have not set out to deal with individual differences in tendency toward imitative behavior, nor have they made explicit statements about the generality of laboratory imitative behavior. They have, however, implied that behavioral habits which are highly and consistently rewarded should be displayed in a great variety of situations. If one is to make useful interpretation of laboratory observations such as those in the present study, one must make some attempt to gauge the extent to which the laboratory behavior is typical of behavior in other situations.

The occurrence of "imitative behavior" in commonplace situations can conceivably be gauged by frequency count by observers provided with some explicit and workable definition of the behavior to be noted. The children in the present study, because of their enrollment in a laboratory preschool, were available for such observation over a period of several months. There are, however, some critical difficulties in defining explicitly a situation of imitative behavior corresponding to the specification of imitative behavior in the laboratory. In order to approximate the conditions of this experimental observation of imitation, a behavioral example must be offered by one individual for potential imitation by the observed individual. Because the frequency with which such examples are offered is not high, and because their occurrence is unpredictable, it would be difficult if not impossible to obtain observational records of such events. It might be assumed, however, that in the absence of offered examples for behavior, the highly imitative (by the laboratory definition) child might overtly seek out information and direction from others.

Method

In connection with other research conducted in the laboratory preschool, an opportunity for obtaining categorical frequency counts of "dependent" behavior arose (Mann, 1959). Two observers were provided with instructional statements defining a number of categories of behavior assumed to demonstrate dependent interaction between the observed *S* and either his peers or his teachers. The observations were executed on the basis of time samples, with a total of 55 two-minute observations on each *S*. The time samples for any given *S* never covered consecutive time spans, and in all cases they occurred within an interval of five weeks. Because the observations of dependent behavior were not executed for all children enrolled in the preschool, observational measures were available for only 21 of the original 32 *S*s.

The total frequency count of "dependent behavioral acts" included five categories: (I) seeking physical contact; (II) seeking proximity; (III) seeking sympathy, protection, or affection; (IV) seeking approval or companionship; and (V) seeking information. While the total count may be interpreted as an index of generalized tendency to behave in a dependent fashion toward others, the concept of dependency employed in this manner is so broad and inclusive that its interpretation is difficult. Category V, however, is relatively specific in its content, and it generally meets the criteria for a commonplace situational approximation of the kind of imitative behavior observed in the laboratory. Within Category V, three subcategories of behavior are described: (*a*) seeking instruction or guidance; (*b*) seeking assistance in decision making; and (*c*) general help seeking.

Interobserver agreement for two observers, graduate students of opposite sex, was estimated by comparison of sampled simultaneous observations of the same child. Considering agreements and disagreements regarding the categorization of individual units of observed behavior, an agreement ratio (agreements/total observations) of .91 was obtained.

Results and Discussion

The product-moment correlation coefficients between observed dependency measures and laboratory measures of imitation are reported in Table 8. The correlations between laboratory imitation and total dependency are greater than those between laboratory imitation and information-seeking dependency, but none of the coefficients is statistically significant. The observational categories as defined here do not appear to relate to the laboratory observations of imitative behavior.

There are without doubt important differences between imitation of peers and imitation of adults. Since the observational measures do not differentiate behavioral units directed toward adults from those directed toward peers, there is a fundamental difference between these measures and the laboratory measures, which deal only with the imitation of adults. While both the laboratory measures and the observational measures probably constitute meaningful data, the absence of any direct relationship between the two should serve to make the researcher cautious in his interpretative generalizations about complex forms of behavior.

SUMMARY AND INTEGRATION

The data presented here should be assimilated critically. There are clear impli-

TABLE 8

CORRELATIONS BETWEEN LABORATORY IMITATION AND OBSERVED PLAY-ACTIVITY DEPENDENCY (Pearson *r*)

Observational Measures	Laboratory Measures	
	Total Imitation	Deviation from Chance Imitation
Total Dependency	−.08	+.37
Information-seeking Dependency (Category V)	−.10	−.15

cations of imprecision in the presently available techniques for assessing parental attitude and practice variables, and there is evidence that generalizations about broadly defined personality phenomena from controlled laboratory observation should be guarded. Furthermore, the number of *S*s employed was unfortunately, but necessarily, small. An effort was made to recognize such methodological difficulties in both the statistical analysis of the data and the interpretation of the findings. There are nevertheless a number of findings which should be valuable in the execution of future research in related areas.

The basic experiment executed here is concerned with training preschool children to imitate the behavioral response of an adult leader. In spite of an initial tendency to make the response opposite to that of the leader in a two-choice discrimination learning problem, learning did occur over a series of 24 trials in which imitation was reinforced consistently regardless of concomitant environmental cues. Older females and younger males were observed to imitate significantly more often than younger females and older males, and a very slight tendency for older *S*s to imitate male leaders more often than female leaders proved to be statistically insignificant.

Utilization of the leader's response as a determinant of behavior (either by consistent imitation or consistent nonimitation) over the entire training series was not appreciably different among the eight subgroups studied. However, while the over-all evidence did not indicate that the training series produced general learning to attend to the leader's response as a social cue, the data suggested that training to imitate a male leader operated differently from training to imitate a female leader. The initial tendency to attend to the female leader's response as a behavioral cue was significantly higher than the same tendency to attend to the male leader's response. Furthermore, while training produced no effect in directing attention to the female leader's response, it did bring about direction of attention to the male leader's response as a behavioral cue.

Although a certain portion of the observed intersubject variance in imitation under prescribed laboratory conditions may be accounted for by the age and sex of the *S* and the sex of the imitated leader, there still remains a considerable range of variance among individuals to be ascribed to other operating variables. An investigation of intellectual capacities indicated that intelligence was probably not a crucial variable in this context.

Exploration of child rearing antecedents revealed a general pattern of relationships to imitative behavior in the child. The techniques employed here for assessing parental attitudes and practices yield highly variable and probably contaminated measures, but there is some consistency among the observed relationships between laboratory behavior and both parental attitudes and reported maternal practices. The predisposition to imitate the adult leader directly was related to such antecedent variables in child rearing as parental strictness or nonpermissiveness, maintenance of parental authority over the child, and stringent control of the child's autonomous independence. A recent study of group conformity in the laboratory by Mussen and Kagan (1958), using TAT type projective assessment of characterizations of parents, indicates that extreme conformists tend to perceive their parents as harsh, punitive, restrictive, and rejecting. In the light of these findings, as well as those of other research (Meyers, 1944; Newell, 1936; Radke, 1946; Symonds, 1939), it is conceivable that this perceptual characterization may not be altogether a projective phenomenon, but may largely be an accurate and realistic perception.

In addition, the predisposition to utilize the leader's behavior as a cue for the child's own behavior was related to a set of variables which may be construed as indicating possessive parental intrusion upon the child's independent activity and exploration. The maternal interview scale of the amount of affectionate interaction with the child (II-23) may possibly index a tendency to "baby" the child and to maintain relatively infantile patterns of relationship with the child. Although it is meaningless to consider correlational indices which may well have arisen by chance, the magnitude of relationship between the deviation from chance imitation measure and certain variables of parent–child interaction suggests that more refined measurement may produce significant findings. Scale III-14 on the maternal interview, which assesses the amount of "play" interaction between the mother and child, correlates +.46 with this measure for boys; Scale III-57, assessing the extent of use of democratic reasoning as a discipline technique, correlates −.45 with the same measure. If one interprets these observations not as conclusive evidence but as worthwhile leads for future research, one is led to the hypothesis that although extensive democratic interaction with the child may bolster his inclination to behave independently, overpossessive "smother love" interaction may interfere with such inclination. An interesting supportive adjunct to such an analysis is provided by the fact that Factor W (Warmth of the mother–child relationship) is correlated —.42 with the deviation from chance imitation scores for the total sample. This coefficient approaches but does not attain the arbitrary .05 significance level (p = .08). An inspection of the constitution of this factor reveals that it includes, in addition to scales of affective warmth toward the child, two scales rather directly related to democratic rationality in dealing with the child: the use of reasoning as a technique of discipline (Scale III-57) and administration of verbal reinforcement for appropriate behavior (Scale II-45). These tentative implications suggest that Sears' Factor W (Warmth of the mother–child relationship) and Factor R (Responsible child rearing orientation) could advantageously be subjected to some purification that would set apart measures of restrictive or intrusive interaction from measures of democratic, equalitarian exchange. The factor analytic approach is not likely to contribute to such purification so long as the cross-contamination occurs within individually defined subscales or at the level of individual response items.

The differences in prediction of behavior for male and female subjects using the child rearing measures employed here are curious. In general, it appears that fairly simple and direct relationships obtain for females but not for males. At least two plausible analyses may account for this discrepancy. It may be that there are meaningful differences between males and females at the stage of development of the present experimental population. While females in this sample have consistently been immersed in a strongly dependent relationship with adults which continues even after the onset of differential training for their eventual adult sex role, the males in the sample are at a stage of development at which differential training for the adult sex role is just beginning. Inasmuch as the dependency-submission issue in social development is one of critical relevance to the differentiation of sex roles, the stage of transition from practically sexless infancy to sex-typed childhood may be a time of some confusion and inconsistency in related areas of behavior. In addition, there is some general evidence of sex differences in the behavior of young children in other areas related to this study. In reviewing a wide range of observations of children's activities, Terman and Tyler (1954) suggest that girls tend to be more "person-oriented" and boys more "thing-oriented." Although there is no data to indicate the age at which such differentiations emerge, it is likely that sex-typed training during the critical period from two to four years of age bears heavily upon the development of such sex differences.

It is particularly interesting that a number of measures related to parental permissiveness were negatively correlated with matched dependent behavior for girls, but positively correlated with such behavior for boys. While none of these inverted correlations for boys were statistically significant, the direction of relationship was consistent for the PARI Factor I scale (Suppression and Interpersonal Distance), the Permissiveness and Punishment for aggression scales on the maternal interview, and the Factor S (Strictness–permissiveness) interview scale. If such relationships were to hold up in future exploration using more refined techniques of measurement, there may be some rather remarkable differences between boys and girls in their response to permissive upbringing.

The observed pattern of relationships between child rearing antecedents and imitative behavior implied by extension that certain differences in imitative tendency as a function of the child's position in the sibling constellation should occur. Experimental data supported the hypothesis that first-born children should be significantly more inclined to imitate the behavior of adults than subsequent children.

An exploration of the generality of the imitative tendencies observed in the laboratory indicated no relationship to a traditional method of assessing dependency in the play activity of children.

While the major contribution of the present findings concerns the importance of two general areas of parent–child interaction (a strictness–permissiveness dimension, and an intrusiveness–overprotectiveness dimension) in determining the tendency for young children to imitate adults, there were other important findings which should offer some guidance in further exploration of the relationships between parent–child interaction and the acquisition of techniques for independent behavior during the course of the child's social development.

REFERENCES

Adler, A. *The education of children.* New York: Greenberg, 1930.

Baldwin, A. L., & Kalhorn, J., & Breese, F. H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58** (3, Whole No. 268).

Champney, H. The variables of parent behavior. *J. abnorm. soc. Psychol.*, 1941, **36**, 525–542.

Chantiny, J. G., Lovell, L. L., & McCandless, B. R. *Parent practices research scale.* Iowa City: State Univer. of Iowa, 1956.

Church, R. M. Transmission of learned behavior between rats. *J. abnorm. soc. Psychol.*, 1957, **54**, 163–165.

Cobb, E. A. Family press variables. *Monogr. soc. res. Child Develpm.*, 1943, **8**, 327–361.

Dean, D. A. The relation of ordinal positions to personality in young children. Unpublished master's thesis, State Univer. of Iowa, 1947.

Gold, R. N. The relationship of two measures of maternal attitudes. Unpublished master's thesis, State Univer. of Iowa, 1958.

Goodenough, F. L., & Leahy, A. M. The effect of certain family relationships upon the development of personality. *J. genet. Psychol.*, 1927, **34**, 45–71.

Grant, E. I. Effect of certain factors in the home environment upon child behavior. Unpublished master's thesis, State Univer. of Iowa, 1937.

Guilford, R. B., & Worcester, D. A. A comparative study of the only and non-only child. *J. genet. Psychol.*, 1930, **38**, 411–426.

Hattwick, B. W. Interrelations between the preschool child's behavior and certain factors in the home. *Child Develpm.*, 1936, **7**, 200–226.

Kawin, E. *Children of the preschool age.* Chicago: Univer. Chicago Press, 1934.

Kelley, T. L. An unbiased correlation measure. *Proc. Nat. Acad. Sci., Wash.*, 1935, **21**, 554–559.

Knight, E. M. A descriptive comparison of markedly aggressive and submissive children. *Smith Coll. Stud. soc. wk.*, 1933, **4**, 168. (Abstract)

Koch, H. L. Some personality correlates of sex, sibling position, and sex of sibling among five- and six-year-old children. *Genet. Psychol. Monogr.*, 1955, **52**, 3–50.

Lasko, J. K. Parent-child relationships: Report from the Fels Research Institute. *Amer. J. Orthopsychiat.*, 1952, **22**, 300–304.

Lasko, J. K. Parent behavior toward first and second children. *Genet. Psychol. Monogr.*, 1954, **49**, 97–137.

Levy, D. M. *Maternal overprotection.* New York: Columbia Univer. Press, 1943.

Mann, N. An investigation of dependency behavior in preschool children in relation to maternal behavior and attitudes. Unpublished master's thesis, State Univer. of Iowa, 1959.

Meyers, C. E. The effect of conflicting authority on the child. *Univ. Ia. Stud. Child Welf.*, 1944, **20**, No. 409, 31–98.

Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Milton, G. A. A factor analytic study of child-rearing behaviors. *Child Develpm.*, 1958, **29**, 381–392.

Mussen, P. H., & Conger, J. J. *Child development and personality.* New York: Harper, 1956.

Mussen, P. H., & Kagan, J. Group conformity and perceptions of parents. *Child Develpm.*, 1958, **29**, 57–60.

Newell, H. W. A further study of maternal rejection. *Amer. J. Orthopsychiat.*, 1936, **6**, 576–607.

Radke, M. J. The relation of parental authority to children's behavior and attitudes. *Univ. Minn. Child Welf. Monogr.*, 1946, No. 22.

Roberts, C. S. Ordinal position and its relation to some aspects of personality. *J. genet. Psychol.*, 1938, **53**, 173–213.

Schaefer, E. S., & Bell, R. Q. Patterns of attitudes toward child-rearing and the family. *J. abnorm. soc. Psychol.*, 1957, **54**, 391–395.

Schaefer, E. S., & Bell, R. Q. Development of a parental attitude research instrument. *Child Develpm.*, 1958, **29**, 339–362.

Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.

Sears, R. R., Maccoby, E., & Levin, H. *Patterns of child-rearing.* Evanston, Ill.: Row, Peterson, 1957.

Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

Shoben, E. J., Jr. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 101–148.

Solomon, R. L., & Coles, M. R. A case of failure of generalization of imitation across drives and across situations. *J. abnorm. soc. Psychol.*, 1954, **49**, 7–13.

Symonds, P. M. *The psychology of parent-child relationships.* New York: Appleton-Century, 1939.

Terman, L. M., & Merrill, M. A. *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence.* Boston: Houghton Mifflin, 1937.

Terman, L. M., & Tyler, L. E. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd ed.) Ch. 17. New York: Wiley, 1954.

Wilson, W. C. Imitation and the learning of incidental cues by preschool children. *Child Develpm.*, 1958, **29**, 393–398.

Zuckerman, M., Ribback, B. B., Monashkin, I., & Norton, J. A., Jr. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, **22**, 165–171.

(Received December 23, 1958)

APPENDIX A

COMPLETE CORRELATIONS BETWEEN PARI SCALES AND LABORATORY IMITATION
(Pearson r)

PARI Scale[a]	Behavioral Measure					
	Total Imitation			Deviation from Chance Imitation		
	Boys	Girls	Total	Boys	Girls	Total
Maternal PARI						
1	+.30	−.01	+.14	+.28	−.27	+.02
2	+.23	−.33	−.01	+.49	−.33	+.21
3	−.19	+.32	+.08	−.13	+.58*	+.17
4	−.31	+.07	−.13	−.33	+.45	+.02
5	+.05	+.24	+.12	+.08	−.02	+.05
6	−.30	−.21	−.21	−.61*	−.30	−.50
7	−.07	+.35	+.13	−.09	+.36	+.13
8	−.13	−.22	−.12	−.42	+.02	−.27
9	−.30	−.10	−.18	−.42	+.35	−.15
10	−.48	+.45	−.05	−.51	+.50	−.13
11	−.30	−.32	−.31	−.33	−.20	−.22
12	−.30	+.52*	+.03	.00	+.57*	+.10
Factor I	+.34	−.04	+.10	+.41	−.51	+.04
Paternal PARI[a]						
1	+.06	−.35	−.17	+.20	−.32	−.02
2	−.14	−.55*	−.37	+.28	−.60*	−.12
3	−.04	+.20	+.08	−.15	+.24	+.08
4	−.08	+.53*	+.14	−.18	+.46	−.01
5	+.06	−.40	−.15	+.42	−.21	+.17
6	−.05	+.30	+.13	−.22	.00	−.11
7	+.18	−.39	−.15	+.52	−.18	+.15
8	−.13	−.13	−.14	−.24	−.15	−.17
9	−.04	+.20	+.08	−.15	+.24	+.08
10	+.28	+.10	+.16	+.79**	+.18	+.30
11	+.57*	−.12	+.15	−.19	+.65*	+.13
12	+.08	+.07	+.07	+.05	+.31	+.09
Factor I	+.07	−.50	−.24	+.38	−.45	−.01

*$p < .05$. **$p < .01$.

[a] See text for names of numbered scales.

APPENDIX B

COMPLETE CORRELATIONS BETWEEN MATERNAL INTERVIEW SCALES AND LABORATORY IMITATION (Pearson *r*)

Interview Scale[a]	Behavioral Measure					
	Total Imitation			Deviation from Chance Imitation		
	Boys	Girls	Total	Boys	Girls	Total
II-23	.00	+.14	+.06	−.37	+.56*	+.21
II-43	−.20	+.33	+.23	−.42	+.26	−.09
II-45	−.23	+.42	+.21	+.09	+.27	+.35
II-53	+.33	−.24	−.15	+.25	−.35	−.13
II-56	+.48	−.12	.00	+.28	−.46	−.09
II-58	+.47	−.22	−.06	+.20	−.56	−.31
II-61	.00	+.41	+.25	−.37	+.06	−.18
II-62	−.28	+.28	+.07	+.27	−.25	−.17
II-66	−.19	+.27	+.06	−.16	−.33	−.37
II-73	−.43	+.34	−.02	.00	+.22	+.14
III-11	+.47	+.10	+.22	−.31	+.40	+.27
III-13	+.02	−.11	−.05	+.06	−.43	−.18
III-14	−.02	+.17	+.03	+.45	+.19	+.23
III-23	+.13	+.24	+.20	−.07	+.22	+.14
III-30	+.12	−.81**	−.57*	+.27	−.40	−.11
III-31	−.15	+.68**	+.35	−.60	+.24	−.11
III-36	+.29	−.07	+.05	+.20	+.23	+.23
III-47	−.30	−.08	−.18	−.45	−.14	−.29
III-57	+.06	+.03	+.04	−.27	−.41	−.31
Factor S	−.41	+.75**	+.25	−.53	+.65*	+.04
Factor W	−.05	−.29	−.22	−.14	−.52	−.42
Factor R	−.21	−.06	−.12	−.32	−.04	−.21

* $p < .05$. ** $p < .01$.

[a] See text for names of scales.

Vol. 73, No. 17 Whole No. 487, 1959

Psychological Monographs: General and Applied

THE VALIDITY OF THE THEMATIC APPERCEPTION TEST IN THE STUDY OF ADOLESCENT PERSONALITY[1]

WILLIAM E. HENRY
University of Chicago

AND JANE FARLEY
Parmly Foundation for Auditory Research, Illinois Institute of Technology

THE STUDY here reported is a modified matching experiment, in which a group of expert clinicians attempts to identify subjects by TAT analyses alone. Blind interpretations of TAT records from 36 adolescents are systematically compared with (similarly blind) interpretations of selected criterion data. These criterion data include objective test scores, summaries of observations and interviews, and Rorschach tests. The interpretations of all of these criterion instruments, as well as the TAT interpretations, were done by clinicians who had not administered the tests, and who had no background information about their subjects other than age and sex. In order to make subsequent comparisons easier, all instruments were interpreted or summarized within the same conceptual framework, previously developed by the entire team of researchers participating in the study.

The experiment follows the general outlines proposed by Cronbach (1948) for projective validation. The method permits comparison of the total projective analysis with other complex data and, at the same time, minimizes judge bias. The TAT analysis is treated as a "prediction," and matched against statements pertaining to a particular "area" of personality function, which are randomly selected from the criterion data of three of the adolescent subjects. Modifications in the Cronbach method—in particular, the use of an analysis of variance design—permit isolation of the influences of many components of the judging process. Statements can be derived about the differential effects of the various behavior "areas," the effects of particular subjects, order and sequence of presentation to the judges, criterion instruments, etc., as well as many of the interaction effects of these factors.

Matching accuracy was computed in two ways, to represent the two tasks required of the judges: that of *identifying* true items about the subject whose TAT prediction was being compared and that of *discriminating* the items taken from some other subject's criterion data. These scores are separately analyzed as "I" (identification) and "D" (discrimination) scores, respectively.

---

[1] This study was made possible by a grant to William E. Henry from the National Institute of Mental Health, and by further assistance provided by the Social Science Research Committee of the University of Chicago. Peter J. Mills of the Parmly Foundation for Auditory Research graciously provided additional funds for the publication costs of this monograph. The monograph is based upon a dissertation submitted by Jane Farley as partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Committee on Human Development of the University of Chicago.

The authors wish to acknowledge their great indebtedness to the psychologists who undertook the arduous judging tasks in either the pilot or the full study: Hedda Bolgar, Mary Grier, Natalie Haimowitz, Gisela Mendel, Harriet B. Moore, Rosalind Cartwright, Johanna Tabin, Sidney J. Levy, Arthur Hartman, Robert McFarland, Irving Leiden; to Eric Klinger who assisted in the statistical analyses and in some aspects of the preparation of material for judging; to Robert Hendrickson who assisted in the preparation of material for judging; to Melvin R. Novick, Association of American Medical Colleges, for his statistical advice, and to David Wallace of the Department of Statistics of the University of Chicago who served as an invaluable statistical advisor throughout the study.

METHODOLOGY OF THE STUDY

The study is essentially a research on individuals, having the particularity and the time span characteristic of a case report. It attempts also the comprehensiveness in depth sampling which is found primarily in intensive examination of a single case; that is, it assesses the full range of personality levels from overt behavior through the directly underlying motivational structure of current needs and press to the basic structural underpinnings of the personality and the dynamic actions and interactions of basic drives and defensive patterns. However, we have attempted also—through itemization of criterion data—to make use of the greater precision inherent in the trait method of validation, while preserving a comprehensive approach to the total personality.

In terms of the particular subgoals served, we have represented most explicitly the "predictive" and the "concurrent" types of validity (American Psychological Association, 1954). The predictive aspect derives from the fact that projective predictions are made from responses to a single TAT administration, analyzed blindly, and compared with criterion data covering several years' duration. The designation "concurrent" refers to the data from other specific tests against which the TAT is matched. Some indication of the "content validity" of the test is also given through the item listing of personality variables matched in the research. Implicit in the process of TAT interpretation, there is also some allusion to the validation of constructs—the constructs outlined by Henry (1956), as those underlying TAT responses.

The Sample

The 36 adolescents were selected from a larger group of 108 cases previously studied in a typical midwestern community.[2] All of the 108 subjects (*S*s) were studied over a period of years, but the records vary in comprehensiveness. These 36 were selected as those having the most complete case records and all of the desired criterion data.

It is notable and germane to the study that this population represents an unusual degree of homogeneity, comprising normal adolescents from the same grade level in a single school, all of whom were born within the same year. They lived in a single small community, were participants in overlapping social groups, were subject to similar local events, and occupied generally the same socioeconomic level.

Logic of Choice of Criterion Data

There exists a generally accepted progression from overtness to covertness in the avowed purposes of various personality study techniques. The "objective" paper-and-pencil tests describe more specific aspects of behavior; the observational and interview methods, with their increased dependence upon psychological inference, are often thought to tap some of the motivational patterns; and the projective techniques are thought by some to reflect more basic structural and dynamic layers of personality. The TAT itself does not fit into this clinical division of labor in any clear-cut fashion. It is sometimes seen as unable to predict overt personality or life history material and it is sometimes thought particularly useful in this regard. The Rorschach has been said to deal with "deeper" layers of personality than the TAT. Sometimes it has been reported that the TAT reflects "content" aspects of personality and the Rorschach "structural" elements. We can hardly hope to answer conclusively the question of what the TAT does and does not measure, but our generally rich and varied data should permit us to shed some light on these issues. Our data combine several features infrequently found in validity studies: each investigator subscribed to the same analytic framework, addressing his analyses to particular areas of function, and yet each did so independently of every other participant. The result is a set of closely comparable statements about the same set of persons. We should thus be able, by a procedure of comparing these statements, to see how the TAT interpretations agree with the more overt instruments, how they agree with the more subjective reports from observation and interview, and how they agree with the estimates of the other projective instrument. With that in mind, we have selected as our criterion data representatives of these three general levels of the overt–covert continuum:

---

[2] These cases were studied as part of a study of adolescent character and personality undertaken by the Committee on Human Development of the University of Chicago and supported by grants from the Lilly Endowment, Inc.

At the *overt level,* the Sociogram and Guess Who test (Havighurst & Taba, 1949), a Family Relations Questionnaire (Brown, Morrison, & Couch, 1947), the Stanford-Binet, Standard Achievement tests, and the Emotional Response test (Havighurst & Neugarten, 1955).

At the *subjective level,* observation and interviews with the *S,* teachers, and parents.

At the *projective level,* the Rorschach.

General Method

To compare these various levels of personality expression with the TAT predictions in each personality area, we have used a design based on a research done by Cronbach in validating the Rorschach (Cronbach, 1948). Not only the general model, but also many of the modifications of that design appearing in this study were suggested or implied in Cronbach's paper. Cronbach was seeking a validational method which permitted a comparison of the unbroken whole of the projective prediction with similarly complex and interrelated data and which did not permit judge bias. He accepted the "blind matching" method as generally satisfying these requirements, but also deplored some of its limitations. The first of these limitations is judgment by elimination (if four sets of statements must be paired, the matching of the fourth pair has little to do with the actual conformity of the statements). Secondly, as Cronbach (1948, p. 367) points out, "If an analysis is partly right and partly wrong, it may be matched with the criterion, and the prediction will seem to be valid; the method does not show the degree of correctness of each prediction, nor the aspects of the prediction that are invalid. . . . " Cronbach's solution was to use the essential method of blind matching, but to force judges to make independent pairings (thereby eliminating judgment by elimination) and to present for matching discrete parts of the prediction, so that only those statements which were correct would be correctly paired with the criterion.

Since, as stated, this design is based on the original Cronbach study, and is more intricate, it is perhaps most easily described by reference to the methods of the earlier research. Cronbach's method, very briefly, is:

1. Clinical predictions (analyses of Rorschach data) were made for each *S.*

2. A criterion was also created for each *S,* in the form of a summary paragraph, independently composed by the *S*'s university advisor (who, presumably, knew the *S* well).

3. The *S*s were divided into groups of three, or "triads." The materials for each triad included the projective predictions and the criterion paragraphs for all three triad members.

a. Projective predictions were broken from summaries into statement items, each statement item being comparable to the statement similarly placed in the statement sheet of every other *S* (Item 1, for example, would always refer to academic problems, though the problems themselves would differ from *S* to *S*).

b. From these three statement sheets, composed from three separate projective predictions, three composite statement sheets were formed, in which items were randomly selected from all members of the triad. (Only one item designated 1 would exist, but might come from the prediction of any of the three triad members, and similarly for all items on each of the composite sheets).

4. A competent judge matched each composite statement sheet with each criterion paragraph for the triad, but no judge saw more than one criterion paragraph for each triad, since this would permit judgment by elimination. Judges rated the statement criterion pairs as either "+," "0," or "?."

5. These judges' decisions were then analyzed to determine the degree of accuracy with which they were able to assign predictive statements to the relevant criterion paragraph. Using only the "+" and "0" markings, Cronbach subjected the resulting 2 × 2 table (categories: (*a*) "+" correct, (*b*) "+" incorrect, (*c*) "0" correct, and (*d*) "0" incorrect) to a chi square analysis. He employed also a measure of correlation (the phi coefficient, because of his dichotomous data) to assess the agreement between prediction and criterion.

In terms of actual scores, Cronbach's results were somewhat disappointing, but these low scores seemed to bear out his contention that many of the previously reported high validity measures, based on blind matching, were misleadingly high. Consequently, he felt that the method, which he admitted was "laborious," had promise in presenting a measure of congruity which was not falsely inflated by spurious partial matching. He was, however, dissatisfied with many of the aspects of his method, and pointed out that the reported research was a trial run of a new technique, leaving much to be desired as a finished design.

Part of the burden of the seemingly poor conformity between prediction and criterion lay with the shortcomings of the criterion used. The descriptive paragraphs were unsystematic, and thus varied greatly in comprehensiveness and depth. In addition, they were based on one individual's viewpoint only—a fact which could introduce the personal bias factor that Cronbach's design had so carefully factored out. Itemization of the projective prediction into separate statements was also, according to Cronbach (1948, p. 368), "a definite limitation of the proposed validation method, since an assessment portrait is more than the sum of its statements. . . ."

Finally, in searching out the weaknesses in the design, as applied, he avowed himself caught between the need for large numbers of judges (necessary for replication of each statement criterion matching) and the need for "hand picking" expert judges. His solution was to forego the hand picking, in order to secure the required number of judges. The design of the current study is an extended and modified version of the study just described. For each of the (12) triads of *S*s, there are not one, but three criteria against which the TAT prediction is to be matched. The number of statement sheets per triad then becomes nine, instead of three. This greatly complicates the problems of constructing composite sheets and of arranging presentation orders of the data to the judges. However, the very complexity and number of the judge decisions required is intended to resolve, for this study, some of the problems which beset Cronbach in his relatively simpler design.

DETAILS OF PROCEDURE IN THE PRESENT STUDY

1. *Our first major change in the research design described above was the decision to "fragment" the criteria rather than the TAT prediction.* Cronbach's own dissatisfaction with an itemized projective report, plus the fact that our more structured criterion material permitted a fairly natural statement breakdown, were the deciding factors in this modification.

2. *Systematization of the criterion data, preparatory to itemizing.* As noted earlier, all of the data collected in the project from which this study draws its material were addressed to a common outline of personality function. This common outline proposes two broad areas of functioning of the individual personality: *Social Interaction* (including peer relationships and family dynamics) and *Characteristics of the Self* (including mental functioning, emotional reactivity and adjustment, and defense mechanisms). Each of the investigators participating in the project had agreed beforehand on the focus of the research. Thus, each tried to fill in as completely as his instrument would allow the subheadings under these two general sectors of personality. The resulting material from all sources was almost ideally parallel in content. This fact made it possible for us to subdivide the personality outline still further, into 27 statement items,[3] without sacrificing (in most cases) the coverage of all our data—prediction and criteria. The reason for dividing criterion data into equivalent statements was to provide statement sheets for each *S* such that the meanings of each item number would be directly comparable on all sheets, i.e., Statement 1 would always represent "degree of participation in peer group," even though different *S*s would provide different specific data ("active participant," "isolate," etc.); similarly, Statement 16 would always refer to "basic emotional attitude," whether the particular comments about an *S* were "active," or "passive," etc.

3. *Itemizing the criteria.* At this point, we had collected for each *S*: a *TAT analysis* (or "prediction"), which was to be used as a total, integrated picture, and also *3 sets of parallel criterion data* (objective, subjective, and projective), which were to be reduced to statement form. Once we had decided on the contents of the 27 statements, our procedure for itemizing the criterion data for each of the 36 *S*s was as follows:

a. Objective data — paper-and-pencil tests: These data almost itemized themselves. Although the material for different *S*s varied in quality and amount, statements were very much to the point, and little choice of material was involved in the itemizing. Each particular test in this objective battery provided statement material for its own area of specialization, and there was no overlap. Peer group statements came from the sociometric analyses, the family area was covered by the Family Relations Questionnaire, standard IQ and Achievement tests filled in the Mental Functioning Area, and the Emotional Response test was our objective assessment of Emotional Adjustment.

---

[3] See Appendix A.

b. Subjective data—observations and interviews: These were the most difficult and time-consuming statement breakdowns to make. There were varying amounts of data for the *S*s, covering an average period of two or three years of study. They were also much less systematically recorded than the other criterion materials, since interviews were conducted permissively, and the observations were recorded verbatim without a fixed advance outline. The initial selection of the 27 statement items from this source was done by the second author, before looking at any of the other criterion data or at the TAT summaries (in order to minimize biased sampling). Since many of the observation–interview summaries ran to 40 or 50 pages in length, any sampling to cover statement areas was bound to be highly selective. Consequently, each full set of observations and interviews, for each *S*, was rechecked by an independent editor[4] against the item samples chosen. This editor was not a psychologist, and it was not his function to comment on the relevance of particular material chosen for the statement it purported to represent. He was instructed merely to see that the sample was properly representative of the total observation and interview material on any particular statement point. If the investigator's original choice for Item 16 (activity/passivity) had been a single statement from the interviews with mother, such as, "is energetic and full of bounce," the editor reviewing the data would not necessarily add any corroborative opinions of neighbors, of the *S* himself, or of others. If, however, the observer (or a teacher) had produced the information that "subject is often sluggish and inactive," this comment would be added to the mother's statement, to properly represent the "balance" of the complete set of data. Thus, the editor cut out repetitious material chosen for the same item and added material which modified, elaborated, or contradicted the selected material.

c. Projective data—the Rorschach test: Each *S*'s Rorschach summary was also broken into 27 statements, reflecting the statement areas of our outline. Since the investigator had already worked with the observation and interview material for the *S*s, this itemization was undertaken by an independent clinician[5] who used the outline of statement contents. By "independent" we mean, that he had not seen the data of the study. Where occasional Rorschach summaries did not cover a particular statement area, the gaps were filled, when possible, by statements from the "sequence analysis" of the same Rorschach protocol.

4. *Separating subjects into triads.* As noted earlier, 36 *S*s were used (chosen from an original population of 108, tested in the adolescent personality project). The 36 *S*s were those whose data were "richest," i.e., for whom all the selected criterion data were available. The 18 males and the 18 females with the data judged most fruitful were chosen. First, the 18 males were randomly assigned to 6 triads, and randomly assigned their positions (A, B, or C) in the triads. The same procedure was followed for the 18 females.

5. *Making up the final composite test sheets.* We had now, for each individual, a TAT summary, and three itemized statement sheets: one objective, one subjective, and one projective. Thus, for each triad, there were three TAT summaries and nine statement sheets. The next task was to combine the items from individual statement sheets into composite statement sheets, which contained materials relating to each of the individuals in the triad.

Starting with Triad I, we began by combining the three "objective" statement sheets (for Individuals A, B, and C) into composite Test Sheets 1, 2, and 3. To do this, we assigned numbers from 1 to 6 to each of the six permutations of the ABC sequence (BAC, ABC, CBA, CAB, ACB, BCA). We then consulted a table of random numbers for random orders of the 1–6 digits. The three composite objective statement sheets were made up according to this formula. That is, if our first three random listings were 1, 3, and 5, this corresponded to the numbered permutations BAC, CBA, and ACB, respectively. This meant that Item 1 would be put into the composite sheets in the order BAC, Item 2 in the order CBA, and Item 3, in the order ACB. In terms of the statements assigned to each composite test sheet, it would mean that Test Sheet 1 would have B's first statement, C's second statement, and A's third statement. Test Sheet 2 would have A's Item 1, B's Statement 2, and C's Statement 3. Test Sheet 3 would have Statements 1, 2, and 3 supplied by C, A, and B, respectively.

This procedure was carried out for all of the 27 items, for each of the three types of criterion, and for each of the 12 triads.

6. *Assembling test materials for presentation to judges.* For each triad, we now had 3 TAT summaries and 9 composite, randomly "scrambled" criterion statement sheets (Statement Sheets 1–3, objective; 4–6, projective; and 7–9, subjective).

---

[4] Mr. Robert Hendrickson, Production Manager for *Today's Health* magazine.

[5] Eric Klinger, Research Associate, Association of American Medical Colleges, and a psychological intern at the Hines VA Hospital, Chicago, Illinois.

To prepare this material for judging, we assembled the statement sheets and TAT predictions into booklets. The introductory pages of each booklet were devoted to the entire TAT summary (prediction). The remainder of the booklet was divided into sections representing the subheadings: Peer Area, Family Dynamics, Mental Functioning, and Emotional Adjustment. The material was arranged so that criterion statements for each area were faced by the corresponding TAT material. These booklets were given to the judges, in a prearranged sequence (see below), together with instructions and an outline of the 27 statement contents.

7. *The judging process.* Nine psychologists, all experienced clinicians and projective analysts, were chosen as judges. This is a considerably smaller number than the 63 judges considered to be absolute minimum by Cronbach. We felt justified in reducing the number of judges, since we permitted each judge to make TAT criterion comparisons on all 36 of the *S*s (Cronbach felt that more than one *S* per triad could not be matched by any judge without bias). To guard against any carry-over of "elimination" effects, no judge in our study compared the same statement sheet with more than one individual; further, comparisons within the same triad were separated by the judging operations for the eleven other triads. We felt that the known effect of "interpolated similar activity," in reducing recall of an original task, would prevent judges from recalling the material for a particular triad across such a large number of decisions on other triads. It was on this basis that we felt justified in our relaxation of Cronbach's stipulations. In a pilot study on only one triad, analysis of variance was unable to show any significant "practice" effect from one prediction statement sheet matching, when a week intervened between this and the subsequent task.

For the nine judges in the main study, three different presentation orders were designed, with three judges (a judge "trio") receiving each of the patterns.[6] By having three judges perform identical tasks, we have some check on the reliability of judging. Although Cronbach specified no less than seven replications of each judging task, the hand picking of judges, plus the very high correlation between decisions of the three judges in the pilot study, appeared adequate justification for this reduction in numbers.

The presentation orders of the matching tasks were balanced, so that each TAT Summary was compared with each test sheet equally often, and so that tasks appearing first, last, and intermediate in the series of total matchings were different for each judge trio. In this way, we hoped to avoid and/or to detect practice effects which might affect decisions made later in the series of matchings. In addition to facilitation with practice, other effects of lateness in the series might stem from the sequence of criteria judged (one criterion matching might sharpen a judge's perception of other criterion information), from a particular sequence of presentation of the individuals in a triad, etc.

Scoring instructions were given to judges in written form, for standardization. Judges' ratings were made on separate scoring forms.[7]

ANALYSIS OF DATA

Two general types of information were sought: first, a *measure of agreement* between the TAT and the criterion data and, second, an evaluation of the *variation in agreement* owing to different matching operations, including some indication of the major sources of such variation. For these two kinds of measurement, we have undertaken two relatively independent analyses.

"Agreement" refers, in this study, to the degree of accuracy with which statements on criterion sheets are assigned to the individual from whose data they were actually taken, guided by the TAT prediction. The Agreement measurement is given by a method suggested for this study by David Wallace, of the Department of Statistics, University of Chicago. The Wallace index serves a purpose analogous to the chi square score in the earlier Cronbach study, that is, it provides a measure of the degree to which the judges' correct ratings exceed the number of correct ratings expectable by chance alone.

This analysis does not rely, as does the chi square, on the assumption of independence of each single matching. The only independence which must be assumed is that between blocks of scores. Even so, it is an assumption we cannot totally fulfill. We have incorporated into the research at each point every possible control against such dependency. To begin with, the data itemized for this research is the generally much more fractionable criterion information. The resulting statements are less dependent on the total pattern than would be statements taken from a projective summary. Even so, we cannot assume that each statement about the *S* is totally independent of every other. Secondly, randomization of the statement items on the composite test sheets, together with balanced presentation order of the material to the judges, renders it difficult, if not altogether impossible, for a judge to make any decisions by elimination. Finally, the heterogeneity of the usual random sample of normal *S*s is decreased in our extraordinarily similar population. This fact implies a decrease in strong contrast between *S*s, and a consequent increase in the applicability of all statements to all *S*s.

---

[6] See Appendix A, Table 3.

[7] See Appendix A, Figure 1.

Despite these efforts at control, it is still possible that a judge's response to one item matching might limit the decisions he could reasonably make on subsequent items. But such qualifications apply to most statistical analyses based on sampling techniques when these methods are used to evaluate organismic, psychological variables.

Our agreement statistic thus shares some of the limitations of the chi square, but has several advantages over that statistic; it is parametric and, therefore, uses the available data more efficiently. Wallace's index also enables us to move close to Cronbach's goal of assessing the degree of agreement, by giving a value of zero to chance matching and a value of +1 for perfect matching. Although the exact nature of the intervals on our agreement scale is unknown and the scale is unstandardized, we do know the lower and upper limits of the scale. If we cannot say for certain how "good" a score of, e.g., .30 is, we can say that it is "better" than a score of .20 and "less good" than one of .40.

In our measure of agreement, two types of error are distinguished and computed separately:

1. *Errors of discrimination:* Because of the homogeneity of the population, there is a large (and uncontrollable) source of statistical "error" in the fact that statements applying to one individual in the group may well apply to at least one other in the same triad. There are many cases of actual resemblance between triad members, so great[8] that a technical "mismatching" of this kind (false positive) need not argue against the accuracy of the TAT analysis.

2. *Errors of identification:* On the other hand, if a judge calls "unlike" a statement which does in fact pertain to the individual whose TAT prediction is presented, such errors (false negative) constitute genuine refutation.

Each of these errors has, we feel, a different significance in answering the questions addressed by the study. Consequently, there are two sets of accuracy measurements (D scores and I scores), based on these two kinds of mismatching. The design of the experiment was constructed to support an analysis of variance. The design properties have been utilized to yield this more refined assessment of the differential effects of different conditions of matching: judge differences, criterion differences, practice effects, subject differences, differences of presentation order, etc. We can also examine some of the possible interaction effects: judge/criterion, judge/area, etc.

Because of the systematic outline of both TAT prediction and criterion data, we can also separate the matchings into areas, looking at, for example, all Peer Area statements as a group or all Family Dynamics statements. It is possible to see in which areas the TAT appears to offer the most accurate analysis and prediction, in which, the least. Here, too, interactions may be most important; in particular, we would expect to find some Area-Instrument (criterion type) interaction, i.e., such questions of overt behavior as "degree of participation in peer activities" or "functioning IQ" are probably more accurately answered by the "objective" criteria than by either observation and interview or Rorschach data. Such directly inferrable, atmospheric items as "general family atmosphere" or "relationship between mother and father" should probably be referred most confidently to the observations and interview reports of trained investigators who actually witnessed the family in action. In the same way, there is no instrument in the criterion battery which can solve the problems of "anxiety" or "level and degree of control" as authoritatively as can the Rorschach.

---

[8] Examples of various degrees of *S* contrast within triads are given in Appendix A.

Thus, if the TAT disagreed with the Rorschach about IQ estimate, but agreed with the standard intelligence test, we would be inclined to weigh the agreement more heavily than the disagreement. If, at the same time, the TAT agreed with the observation and interview data on family dynamics, but disagreed with the Family Relations Questionnaire on the same area, again, we could feel fairly confident of the accuracy of the TAT prediction. Finally, if the TAT predictions of intrapersonal dynamics were confirmed by the Rorschach, but at odds with the same material from the observation–interview reports, we would be inclined to take all of these predictive successes more seriously than the failures.

HYPOTHESES, AND EVIDENCE ACCEPTED AS CONFIRMATORY

Having now discussed the properties of the criterion data, the process of comparing criteria with TAT predictions, and the general nature of the proposed statistical analysis, we can more meaningfully state and elucidate the hypotheses underlying our approach. They are as follows:

1. *The conclusions reached by the interpreter in the analysis of a subject's (TAT) responses are "valid."* As confirmation of the validity of the TAT, we will accept measurements of agreement with the criterion which are significantly better than chance at the .01 confidence level or better. (The choice of this level of confidence is, we feel, necessitated by the large number of decisions involved, which can inflate moderate departures from chance to significance.) Where there are differences in the amount of agreement with the TAT prediction among the several criterion instruments, we take as our standard the criteria cited above. These are particularly qualified to assess each of the various areas, namely:

Area	*Standard Criterion Instrument*
Peer Group and Mental Functioning	Objective Battery
Family Dynamics	Observation and Interview Data
Emotional Adjustment	Rorschach

In any cases of divergence among objective, subjective, and projective criterion sheets, agreement with these criteria in the areas assigned to each will constitute acceptable validation of the TAT prediction for that area, even if the other two criterion instruments disagree. By the same token, if the TAT agrees with some other criterion, but does not significantly agree with the one specified for the area, the TAT prediction will be considered to be not confirmed.

2. *The interpretations made by the method herein outlined are reliable.* In the physical experiment, "reliability" is the measurement of reproducibility —the extent to which a given test procedure and instrumentation measures the same thing, every time, for every *S*. Establishing reliability is a matter of calibrating instruments and checking the error function of the tester, since the tested properties do not change.

In psychological testing, the case is somewhat different. Objective psychological tests do permit a relatively simple operational concept of reliability. Because the items of such tests are assumed not to be sequentially dependent, it is possible to compare parts of the same test protocol to see if they say the same thing about the *S*. It is possible also to retest, to see if later administrations of the test agree with the first. But at this point, even in objective testing, problems are raised by the timing element. If the interval between readministrations is too short, the second test may simply be a recall of responses made in the first. If the interval is too long, significant changes may have taken place, which appear (and should appear) in the test. The standard objective measures of, e.g., IQ, take this into account by prescribing the amount and the direction of change. That is, the same relationship of subject-to-test, in terms of score, is represented, at specified time intervals, by specified differences in the actual test performance. But the problem is not limited to objectively measured psychological functions. A valid test, whether objective or otherwise, must take into account both the constancy of the subject-test relationship and the changes in actual responses. These are the familiar Scylla and Charybdis of psychological testing: *stability* and *sensitivity*.

With projective testing, the case is even more complex. Since projective tests do not have a true criterion of "correctness," the dimensions of interpretation are greatly multiplied, or possibly infinite. In those psychometric tests that ignore patterning of responses in arriving at their principal assessment measure, the incorrect responses are, as Rosenzweig (1949), says "thrown away." In projective tests, the "incorrect" (divergent from normal or popular) responses are positive data. "Error" becomes "individuality."

In projective testing, subject-test relationships are multidimensional. Furthermore, the dimensions are highly interdependent, in the sense that interpretation of each one depends on the context of all of the others. Prescribing specific response changes which will define a constant relationship of *S* to test over time becomes prohibitively difficult. Because we are interested in both of the reciprocal functions of stability and sensitivity, the question of reliability becomes increasingly difficult to answer. We ask "How does the subject express his constant self in different contexts" (of age, situation, etc.), and we ask the same test responses to reflect both parts of the answer.

There are, in fact, two general notions of reliability which are often used interchangeably in psychological testing. The first applies to the consistency of the test items themselves; the second, to the efficiency of the test in registering the consistency of an *S* from time to time or from situation to situation. The question of item consistency (as measured by split half, alternative form, or other methods which test homogeneity of test parts) is chiefly of importance in psychometric testing, in which the treatment of scores as additive and evenly weighted requires this concept. The interpretation of projective responses does not depend on such an assumption. Indeed, the establishing of item correspondence in projective tests would destroy one of the major projective test variables: sequence—that part of the stimulus value of a test item which is due to its placement in the series, and its subsequentness to particular other items. In order to utilize the information given by sequence, the projective tester makes it impossible to establish any clear measure of the "equivalence" in either split half or equivalent form tests of item consistency.

It is in the second general type of reliability—stability across time—that projective testers are primarily interested. As noted above, the measuring of stability in projective test responses is complicated by the same factors which posed problems in the stability of objective test performance. There is still the possibility, as in objective testing, of contamination in retesting by a superficial recital of previous responses. Again, if the interval between tests is too long, there is the problem of registering actual change in the *S*, and the changes which may be registered are much more difficult to specify and interpret.

There is yet a third complication in considering the concept of reliability in projective testing; it is the necessity of using test score concepts and methods to analyze data which are composed, not of test scores, but of interpretive opinions. The Rorschach seemingly avoids this difficulty by translating many classes of interpretive opinion into score form, but it encounters the problem face to face, after the scoring is done, in the question of score interpretation. However, researchers have found some compensation for the additional complexity of the interpretation; the recording of interpretive judgments at last gives the researcher some data which can be quantified and compared. If several judges interpret the same test protocol in the same way, then both test and interpretive method are likely to be "reliable." Some clinicians believe that the only meaningful reliability measurement is interinterpreter agreement. Kogan and Hunt (1950), in discussing the reliability problem, reason that the agreement of judges corresponds to the agreement of scores. They reach this conclusion because the judge's analysis is the raw data (corresponding to the data of test scores). Using a single variable (movement of the patient) in casework and a simple interpretive convention (numerical rating of —2 to +4), they use several statistical analyses to demonstrate probabilities of from .001 to .05 for the chance occurrence of the measured agreement between their judges.

There are several major obstacles to the interjudge agreement concept of reliability. The first is the inability to assess particular segments of the entire stimulus-response-interpreter-analytic method sequence. If the judges do agree, all of these processes are likely to be reliable, but if they do not, the meaning is not so clear. The lack of conformity may result from semantic divergence between the judges' understanding of the terms used or from divergent or inadequate interpretive methods. One way around this is to measure the "self-congruity" of a single judge, as suggested by Combs (1947). However, as discussed above, the difficulties in setting time intervals with relation to *S* responses now contaminates the judges' interpretations. A rejudgment, as well as a retest, is subject to the influence of recall of previous judgment. Similarly, if the interval is too long, the judge himself may have changed, or at least changed his interpretive methods. In any case, the interjudge agreement measure is much nearer to "item reliability" than it is to the "stability" concept which more properly pertains to projective testing.

The stability of a projective prediction is possible through confirmation of future predictions made from the test analysis or through recognizability of the *S* by his test interpretation, at points in time previous or subsequent to the time of testing. Both of these factors are built into our study. Because of the longitudinal design of the original project, interpreters of several of the tests utilized the opportunity to make predictions from the test material, to be confirmed or refuted by subsequent events. The TAT is one of the tests containing a number of developmental predictions for most *S*s. Furthermore, the TAT represents a single point in the span of a several-year study, during a period of great change in the lives of the *S*s. It is compared with data which cover the entire span of the research. This applies particularly to the observation and interview material which was continuous throughout the duration of the project, but it is also applicable to much of the other data. Sociometric tests and several of the other objective criteria were readministered at least once, after a several-year interval. In over a third of these *S*s, the Rorschach was also readministered. Consequently, we will accept the reliability (stability) of the test predictions as confirmed if the I and D scores are correctly judged at the .01 level of significance, or better.

It is perfectly reasonable to question the usefulness of any reliability assessment, in experiments resulting in high validity measures, since the function of reliability is to permit valid measurement. In the strictest sense, a reliability test is only necessary if we wish to know how much we can increase the measured validity, since reliability sets limits on how "valid" the test may show itself to be. In this type of validity test, for which the concept of reliability is nebulous at best, we cannot "think of reliability in the usual sense, as the 'sine qua non' of our acceptance of findings" (Combs, 1947, p. 263). But it is our belief that the stability factor, at least, is manifestly worth knowing, and worth testing, since the clinical use of the TAT, like any other diagnostic test, is based in part on the knowledge of its ability to tap enduring characteristics of the *S*'s personality.

3. *The inferences drawn about a given personality from TAT protocols alone are uniquely individual, independent of circumstantial factors.*

a. They typify an individual as an integrated person, who is recognizable from age to age and from situation to situation.

Recognition of *S*s as whole and consistent individuals will be considered to be confirmed by a greater than chance agreement of TAT predictions and criteria, based on the "Identification" scores, and significant at or above the .01 level.

b. Inferences from the TAT represent the person at a deeper level than the socially determined group responses of his milieu.

We will accept as confirmation of this uniqueness factor, a measure of agreement at or above the .01 significance level, based on the "Discrimination" scores. Considering the homogeneity of the population, the ability to separate the individual members of the triad, by TAT prediction, bespeaks the TAT's ability to represent the uniqueness of an individual, as distinct from his social group.

4. *Current behavior, underlying dynamics, salient historical facts, and probable direction of change of an individual can be inferred dependably from the TAT.* Of this hypothesis, we can expect only partial correlation. No one really expects one clinical technique to take the place of a whole battery of others. There may be, for example, areas, or at least items, in which the TAT taps underlying motives only, and is thus "inaccurate" in predicting behavior, as compared to the objective tests in the battery. It may also be that there are parts of the inner dynamics of personality which are "too deep," as some clinicians have suggested, for the TAT to tap. However, we should be able to conclude with some particularity exactly where these areas are, and to provide some explanation of *why* they are. The evidence we will accept as confirmatory of the parts of this hypothesis are:

a. Agreement with objective data (current behavior)

b. Agreement with projective data (underlying dynamics)

c. Agreement with subjective data (salient historical facts)

d. Accuracy over the time span of the study (probable direction of change)

It will be noted that the analysis of these data is set up according to a "null hypothesis" philosophy of significance. This is not, to be sure, the most powerful use of experimental design or statistical method. But the alternative to this usage is the use of experimental and statistical techniques to estimate parameters, that is, to measure the confidence which an experimenter may have in a prediction of outcome which he has already made. This kind of "validation" is appropriate for special fields, or whole scientific terrains, which have advanced far enough to generate such verifiable estimates. Projective testing, as a science, is new enough so that the less rigorous, systematic, exploration kind of research based on a "null hypothesis" concept is still the order of the day. It is our hope that this study may help in the formulation of definitive theoretical statements, which may rightly be tested by the confidence interval method. Indeed, it is possible that subsequent research on these same data may adopt such a form, once our exploratory results are available for direction.

ANALYSIS OF RESULTS

Statistical Procedures

The raw data were in the form of judges' ratings, recorded on the rating sheets which were supplied to the judges, along with the comparison materials. To compute the basic scores used in the analysis, these judge ratings were checked against the correct ratings; each of the 26,244 judgments was then clerically coded into one of 7 possible categories:

1. L+ (Correct rating = Like; judge rated "+")
2. L− (Correct rating = Like; judge rated "−")
3. U+ (Correct rating = Unlike; judge rated "+")
4. U− (Correct rating = Unlike; judge rated "−")
5. L? (Correct rating = Like; judge rated "?")
6. U? (Correct rating = Unlike; judge rated "?")
7. ?? (No criterion information given; judge rated "?")

Approximately 86% of these judgments were coded into the first four categories, and approximately 14% into the last three categories. However, of the judgments rated "?" by the judges, more than 90% were in Category 7, items for which no information was available in the criterion sheets. Inspection of the remaining 357 "?" items showed no rating trends on the part of judges, but items were scattered through the rating sheets in seemingly random fashion. Thus, no formal analysis was made of the doubtful items.[9] The judgments statistically analyzed were those in Categories 1 through 4: L+, L−, U+, and U−. Categories 1 and 4 (L+, U−) are correct judgments, of the *Identification* and *Discrimination* types, respectively. Categories 2 and 3 (L−, U+) are respectively *Errors of Identification* and *Errors of Discrimination.* Following this coding operation, the data were sorted into various "blocks" of judgments, on the basis of the several test variables. Block Scores were computed:

1. For the permutations of 36 *S*s
2. For 12 triads
3. For the four behavior areas
4. For the three test-sheet types (objective, subjective, projective) or "instruments"
5. For the three judge trios (i.e., for each set of judges who performed identical tasks)
6. These data were analyzed in such a way that their block scores could be regrouped and analyzed also for each of the three task orders (first $\frac{1}{3}$ of total judgments, second $\frac{1}{3}$, and third $\frac{1}{3}$) for all nine judges
7. A preliminary exploration was made of the effect of the "sequence" of instruments against which the TAT prediction was matched (i.e., judge matches objective first, then subjective, then projective, or any other of the possible orders). It was felt that, should "sequence" appear to have a significant influence on judges' decisions, it could be factored, balanced, and analyzed
8. The data were also sorted and analyzed for those of the interaction effects which were translatable into meaningful psychological variables

The statistical analyses were performed on the scores of these several "blocks." Each "block score," based on either the Identification or Discrimination decisions (I scores or D scores), represents the percentage of correct ratings in excess of the percentage which would be expectable by chance alone. The rationale for the formulae used in arriving at these scores is as follows:

1. We have separated the Like and Unlike statements, since we felt that different problems were involved in both correct and incorrect ratings of the two. The I scores (both correct and erroneous) are concerned with Like statements (those for which "+" is the correct answer), and the D scores with Unlike statements (those properly scored "−").

2. The Null Hypothesis would state: The "+"s assigned by judges are randomly distributed, and have only a chance relationship wth the Like statements in this block. We can express the proportion of correct judgments as: $\frac{L+}{\text{all}L}$. If the "+"s are not systematically related to the Like statements, then the expectation of correct judgments of Like statements would be no greater than the proportion of "+"s made on all items, i.e., no greater proportion of "+"s will be assigned by chance to the Like statements than to the same number of statements chosen at random from the total. Then the degree to which the actual correct L+ markings exceed the chance value would be: $\frac{L+}{L} = \frac{+}{L+U}$ with chance equal to zero. By the same rationale, the formula for arriving at the percentage of correct ratings, above chance, in the D score would be: $\frac{U-}{U} = \frac{-}{L+U}$.[10]

Making use of the fact that chance alone would give a value of 0 to these scores, it is possible to define a confidence level for the significance of the different-from-chance values of the scores. The index,[11] suggested by Wallace, for determining the significance of the judgments, based on the null hypothesis of chance distribution is: $F = \frac{N(X_G)^2}{S^2M}$, where X_G = the grand mean, and S^2M = the mean square of trios plus the mean square of triads, minus the trio-triad interaction. This denominator was used, rather than the more conventional residual variance, in order to minimize the interdependence of the separate judgments which comprise the blocks. In particular, this formula takes into account any systematic effects of the judges'

---

[9] A summary table of the "No Information" items (Table 7) will be found in Appendix B.

[10] This formula could also be expressed as a *t* value:

$$F = \frac{N(X_G)^2}{S^2M} = F = \frac{\frac{X_G{}^{-2}}{S^2M}}{N} = t = \frac{\frac{X_G{}^{-0}}{SM}}{\sqrt{N}}$$

[11] It must be noted that this formula is occasionally misleading. Because of the random assignment of *S* statements to the composite test sheets against which TAT predictions were matched, it is possible for any area of a test sheet to contain statements about a *S* in widely varying proportions. It is possible for a judge to compare the TAT prediction with a criterion area in which all of the statements were, e.g., unlike the *S*, because they were taken from the criterion data of the other individuals in the triad. A judge making a totally correct set of judgments in such a case $\frac{U-}{U} - \frac{-}{L+U}$ would come up with a score of zero for that block of statements, since both ratios would = 1. However, this is only true in the few extreme cases, and since statement assignments were random, this is assumed to be a random contamination, which will not materially affect comparisons between the scores.

individual personalities, and the nonrelevant differential effects of the subject-personalities involved in the various triads. The degrees of freedom figure for the denominator is found by the formula:

$$\frac{1}{df}\text{(of denominator)} = \frac{1}{df\ x}\left(\frac{x}{x+y+x}\right)^2 + \frac{1}{df\ y}\left(\frac{y}{x+y+z}\right)^2 + \frac{1}{df\ z}\left(\frac{z}{x+y+z}\right)^2$$

where x = mean square of trios, y = mean square of triads, and z = mean square of the trio-triad interaction.

RESULTS

Results of the analyses of variance, and the validity index for these scores are summarized in the two tables below: (The F values marked "+" are significant at .05–.01 level. Those marked "++" are significant at and above the .01 level. All others are nonsignificant.) Details of computation, and block scores for each of the blocks will be found in Appendix B.

TABLE 1

I-SCORES

Summary of Analysis of Variance

Over-all Significance: $F = \frac{N(X)^{-2}}{S^2m} = 64.026^{a}(.01 = 21.20)$

	df	*ss*	*ms*	*F* (++ significant at .01 level or better)	Denominators Used for *F* (mean square of effect No.)	*F* Required for Significance at .01 Level
Effects						
Total	3887	4,848,488.				
Between Trio	2	31,778.	15,889.000	8.829++	13	5.72
Between judge within Trio	6	58,074.	9,679.000	12.506++	14	3.12
Between Triad	11	103,725.	9,429.545	9.152++	14	2.82
Between subject within Triad	24	118,817.	4,950.708	5.003++	20	1.79
Between instrument	2	2,510.	1,255.000	.391	15	5.18
Behavior area	3	135,829	45,276.333	19.712++	16	4.13
Order	2	12,283.	6,141.500	6.206++	20	4.61
Interactions						
Total		1,062,667				
Test Sheet–Area	6	10,679.	1,779.815	1.798	20	2.80
Trio–Area	6	16,917.	2,819.449	1.228	16	3.12
Triad–Area	33	66,988	2,029.939	.884	16	1.60
Instrument–Trio	4	17,880	4,470.110	1.391	15	3.83
Instrument–Triad	22	30,480	1,385.449	1.409	15	2.29
Trio–Triad	22	39,592	1,799.646	2.325++	14	2.12
Judge (within Trio)–Triad	66	51,081.	773.954	.782	20	1.47
Trio–Triad–Instrument	44	141,379.	3,213.167	3.247++	20	1.59
Trio–Triad–Area	66	151,593.	2,296.865	2.321++	20	1.47
Trio–Instrument Area	12	14,966.	1,247.175	1.269	20	2.18
Triad–Instrument Area	66	141,388.	2,142.250	2.164++	20	1.47
Trio–Triad–Instrument Area	132	430,763.	3,263.361	3.297++	20	1.32
Residual	3358	3,322,805.	989.519			

[a] This F value is given by the denominator outlined on page 32. A less conservative estimate of F, based on the Trio–Triad Interaction alone, as a denominator, yields an F value of 836.953.

TABLE 2

D-Scores

Summary of Analysis of Variance

Over-all Significance: $F = \frac{N(X)^{-2}}{Sm^2} = 98.796^a$ (.01 = 34.12)

	df	*ss*	*ms*	*F* (+ = signif. @ .05-.01 ++ = signif. @ .01 or better)	Denominators Used for *F* (mean square of effect No.)	*F* Required for Significance at .01 Level
Effects						
Total	3887	1,431,552.				
Between Trio	2	2,738	1,369.000	1.479	13	5.72
Between judge within Trio	6	22,534	3,755.667	12.471++	14	3.12
Between Triad	11	13,284	1,207.636	2.971++	14	2.82
Between subject within Triad	24	22,746	947.750	3.134++	20	1.79
Between instrument	2	765	382.500	.265	15	5.18
Behavior Area	3	10,287	3,429.000	11.341++	16	4.13
Order	2	2,962	1,481.000	4.898++	20	4.61
Interactions						
Total		340,979				
Instrument Area	6	1,393.	232.289	.768	20	2.80
Trio–Area	6	4,867.	811.302	1.079	16	3.12
Triad–Area	33	20,285	614.720	.818	16	1.60
Instrument–Trio	4	1,415.	353.793	.245	15	3.83
Instrument–Triad	22	10,544.	479.317	.332	15	2.29
Trio–Triad	22	20,359.	925.422	3.963++	14	2.12
Judge (within Trio)–Triad	66	15,412	233.561	.772	20	1.47
Trio–Triad–Instrument	44	63,471	1,442.541	4.771++	20	1.59
Trio–Triad–Area	66	49,607	751.632	2.486++	20	1.47
Trio–Instrument Area	12	6,604	550.374	1.820++	20	2.18
Triad–Instrument Area	66	37,087	561.924	1.858++	20	1.47
Trio–Triad–Instrument Area	132	125,347	949.554	3.140++	20	1.32
Residual	3358	1,015,257	302.339			

[a] The trio-triad denominator yields an *F* for D scores of 749.107.

The interjudge reliability was estimated by selecting pairs of judges from each trio and comparing their performances in terms of frequency of occurrence in each of the 4 coded categories (L+, L−, U+, U−), by means of 2 × 2 contingency tables and the formula

$$\text{Phi} = \frac{AD - BC}{(A+B)(C+D)(B+D)(A+C)}, \text{ in}$$

which the dichotomous values of 0 (negative judgment) and 1 (positive judgment) represented each judge decision. Phi values for the three trios were: For Trio 1, .414; Trio 2, .380; and Trio 3, .459. Although phi tends to yield somewhat lower scores than the corresponding rho values, none of these pairings can be assumed to correlate above the .5 level.

INTERPRETATION OF RESULTS

Consequences for Specific Hypotheses

1. *The validity of the TAT.* We selected, as the evidence for confirming our validity hypothesis, I scores which showed a greater than chance relationship between TAT and criteria, at the .01 level of confidence. The over-all analysis of I score blocks shows significant agreement well beyond the agreed-on level, and we may consider our first hypothesis to be confirmed. There was no significant variation in the agreement of each of the criteria with the TAT Prediction.[12] Consequently, it was not necessary to base the confirmation of this hypothesis upon the somewhat arbitrary selection of a "standard" criterion for each area.

Despite the rather surprising lack of conformity among the individual judgments of these expert judges, the combination of judges' scores matched the correct "I" ratings at or above the .01 level of confidence. We will have more to say about the meanings of these effects below, but it is safe to assume that the lack of judge agreement reduces the apparent TAT criterion measurement and that the actual agreement between our projective prediction and the criteria used may be even higher.

2. *The reliability of the TAT.* As predicted, the "stability" aspect of reliability was the only one on which we are able to reach any conclusion. The fact that both I and D scores exceeded expectable chance values for all *S*s allows us to say that the TAT predictions have withstood the test of time. The criterion data matched against the TAT were based in many cases on several years' study. The TAT interpretation was drawn from a single administration at some point during this interval of study. The agreement of the TAT with these long-term observations and with the repeated criterion testings reflects the value of the test in predicting future trends and inferring at least the recent history of these *S*s.

3. *The discrimination power of the TAT.* We find that our hypothesis about the usefulness of the TAT in expressing the uniqueness of individual personality is confirmed by the high D scores. The measure of conformity on these scores indicates that the TAT prediction provides a guide for a trained clinician, by which he is able to recognize a unique individual, even against a social background which equates virtually every aspect of that individual except his individuality.

4. *The scope of the TAT.* The instruments which provided our criteria for comparisons were assumed to represent the different personality levels. The lack of significant instrument variation in TAT criterion agreement, together with the stability of prediction discussed above, allow us to claim for the TAT, within the limits of our population and our analytic tools, the wide scope which our hypothesis originally stated: all three levels of behavior, and the past and future dimensions of time as well as the present, can be dependably inferred from TAT data alone. However, some of the differences which were anticipated in the first formulation of this hypothesis were found. The TAT does differ in its predictive efficiency as a function of the area of prediction, the individual *S*, the judge making the decisions, etc. Although none of these factors "disqualifies" the TAT prediction, all of them provide us with increased understanding of the differential sensitivity of the TAT in various contexts.

Discussion of the Major Findings

1. *The high significance level* of the TAT criterion agreement in all contexts is certainly our principal concern. However, there are several qualifications which might be attached to this finding:

a. The large number of judgments tabulated in computing scores: It is clearly true that the larger the number of matchings used, the smaller the magnitude of consistent agreement necessary for a significant level of congruity. This "inflating" property of the large number of matchings analyzed is, we feel, more than compensated by the difficult task we have set for the TAT. The "blind" analysis of the *S*'s responses was undertaken on a group which was homogeneous for virtually all variables except unique personality. Then this analysis of a single TAT administration was matched against data intensively collected over a several-year period. The personality information with which TAT predictions were compared was comprehensive in level and scope and represented a duration of time which was—with respect to the single TAT administration—both past and future, as well as concurrent.

---

[12] See full discussion of this finding below.

b. The low interjudge reliability: The lack of conformity shown between judge's matchings was not anticipated. The finding is not unprecedented. Interjudge reliability for similar tasks has been reported variously as high (Krugman, 1942), very low (Cox & Sargent, 1950), and indeterminate (Havighurst & Taba, 1949). However, we had expected much higher correlation on the basis of our judge selection and the results of the pilot research. Pairings of the three judges in this pilot study yielded reliability coefficients of .87, .91, and .94. Two of the same judges participated in this study, and the other seven were equally trained and similarly oriented. The only known difference in the selection of judges between the pilot study and the larger study was that all three judges of the smaller research were women, while the present judges are five women and four men. There is no obvious relationship between gender and interjudge agreement, and since the single pair of female judges whose agreement was computed produced essentially the same level of agreement as the other two pairs, it is probably safe to assume that the gender of judges was not the determining factor in producing these unexpected results.

The changes made in the tasks required of the judges from the pilot study to the larger research are far more likely to have caused the difference in judge agreement. These changes were of two kinds: a change in the number of matchings which the judges were asked to do and a change in the actual material presented—different individuals, different triads (implying different levels of contrast), different criterion sheets. Our analysis of variance strengthens the suspicion of the influence of a change of task. Variations in the judging process due to the effects of different triads, different individuals, etc. reflect the changes in performance due to these factors. The number of tasks seems also to have influenced the quality of the matchings—judged by the "fatigue factor" shown in the significant "order" effect (see below). But whatever the reason,[13] the low interjudge reliability must affect our confidence in the high validity of the TAT predictions. The actual congruity between predictions and criteria is not called into question, since the low judge conformity must reduce, not enhance, the apparent validity. It would therefore produce a low estimate of actual validity. However, the divergence between judges might well cause us to hesitate in asserting the same degree of predictive efficiency for the TAT when used in the clinical situation. We are forced to say: "The validity is there, *if* the clinician can make use of it."

c. Failure of the data to meet the assumptions of the analytic methods: We have mentioned a single distortion which might result from our statistical techniques —the false "chance" score which could result when the statements in a given block are either all "like" or all "unlike" the *S* against whom they are being matched. Since the assignment of statements to test sheets was random, this effect might lower the over-all validity estimate, but would only lower it slightly. We would not expect it to have any systematic distorting effect, i.e., on any particular blocks of scores. Consequently, it would not alter the variance analysis.

There is, however, another aspect of our techniques which might be more seriously questioned; this is the fact that the *F* test assumes normally distributed, equally variable, and independent scores. The "independence" we have tried to enforce. While no datum concerning a single individual is ever totally independent of other datum which refers to the same individual, we have described the precautions taken (in processing criterion information) against the giving of artifactual "clues," which might link the separate statements (i.e., idiosyncratic verbal habits of interviewed parents). An attempt was also made in the computing of significance of score to control against the remaining dependency factors. This was described in the preceding section. It is our belief that these statements are now virtually independent, except for the inherent resemblance of true statements about a single *S*.[14] The assumption of normality is obviously not met by the distributions of either I or D scores. Inspection of the frequency histograms of I and D scores[15] reveals their nonnormality (the D distribution is approximately symmetrical, but leptocurtic; the I distribution has a considerable negative skew). However, it has been demonstrated empirically that neither nonnormality nor heterogeneity of variance—nor a combination of the two—vitally affects the accuracy of the *F* ratio as a test of the null hypothesis concerning the means of two data samples.[16]

---

[13] Questionnaires were completed by all judges, with respect to their work patterns, clinical frame of reference, amount and variation of interest in the tasks, etc. A later test is planned to investigate further the effect of these personal variables on the judges' performances.

2. *Differences between I scores and D scores.* It is of interest to see whether our division of the matching tasks into the two processes of identification and discrimination agrees with the results obtained for these two sets of scores. Our first observation is that both scores were very significantly higher than chance. A closer look reveals the fact that the actual numerical value of the *F* ratio for over-all significance is greater in the case of the D scores than the I scores. This appears to be contrary to our original contention that the scores should be separately computed, because the task of discrimination, in a population as homogeneous as this one, would be a much more difficult one than that of identification. Errors of the D variety we felt to be easier to make, and in a sense justified by the data. The difference in values of the *F* ratios seems to say that exactly the reverse is true. However, a glance at the *ss* columns of the two summary tables offers a clue to the seeming disparity. The variation in TAT-criterion agreement based on the I score analysis is consistently greater than the same measure using the D scores. The lower variability of D scores reduces the denominator of the *F* ratio and thereby increases its value. The grand mean for I scores is, as expected, larger (19.680) than the D grand mean (13.353).

The lower variability of D scores can be at least partially explained by the greater number of D decisions made. For each TAT prediction, only about $\frac{1}{8}$ of the statements with which it was matched were actually Like statements. There is also a possibility that some of the difference in variability (between I and D decisions) can be attributed to the different nature of the tasks. The I judgment—deciding whether a correspondence does exist between an *S*'s TAT prediction and a given statement—was expected to be easier and more obvious and superficial than the exacting, hair-splitting decision that a statement does *not* match the prediction (D judgment) in a homogeneous population. It is conceivable that the I judgments tended to be more spontaneous, to involve less intense concentration, and perhaps to reflect more the interpersonal differences between judges than the D judgments. The much lower "between trio" and "between judge (within trio)" mean squares on the D scores might tend to bear this out. Thus the D scores may show lower judge variability because the judges, in deliberated decisions, behaved more like clinicians, while in the spontaneous I judgments, they behaved more as persons. However, the I and D scores show generally parallel trends, and the discussion to follow, on the principal effects noted in the analysis of variance, can thus be applied justifiably to the combined I and D findings for each variable.

---

[14] We are partially corroborated in this belief by the fact that *F* scores can be obtained using the residual mean square as a denominator which are very similar, in every case, to the values based on the special denominators.

[15] See Appendix B.

[16] (Norton 1952, p. 714) "Even [in the] case with populations differing marked in form and variance, only about 10% of the mean square ratios exceeded the *F* values by 5 percent-points."

3. *The principal effects noted in the analysis of variance.*

a. Area. By far the greatest single effect relates to differences in the behavior area in which a prediction was made (Peer Group, Family, Mental Functioning, Emotional Adjustment). This effect is entirely in keeping with our expectations.

The reason commonly given for assessing personality by a battery of tests rather than a single instrument is that the tests will be focused on different particular areas of personality function. The TAT must also have its particular spheres of maximum efficiency. Scores in the four areas divide themselves neatly into two groups, Area 1 and Area 4 (Peer Group and Emotional Adjustment) of virtually equal percentage above chance. These are noticeably more accurately matched than Areas 2 and 3 (which have also virtually the same score).[17] Since there are no significant effects attributable to either the instrument used (test sheet type) or interaction between area and instrument, we can only discuss the total area effect, as applying equally to all the three criterion types.

Literally, this tells us that the TAT is a better predictive guide to personality function in the context of peer group relations and emotional adjustment (or intrapersonal dynamics) than in the family relations or mental functioning contexts. The pairing of areas in these two divisions has some elements of incongruity. The peer area reflects the most overt of the four areas and, also, the most "social" factor. The other social context—the family area—is considerably involved in matters of covert, emotional currents, and much more focused on the inner significance, to the *S*, of the corresponding behavioral events. The emotional adjustment section, on the other hand, is the least overt and "social" of the four areas; it is almost entirely occupied with the private world of the *S*, conscious and unconscious, and with the relationship of the *S* to himself. It would have seemed more directly logical, on the basis of both the overt/covert and the social/private dimensions, for the "peer group" and "family dynamics" areas to have become paired, pairing also mental functioning and emotional adjustment. The single quality which might fit the division made by the scores is their placement in time. The peer group and emotional adjustment areas are perhaps oriented somewhat more to a current point of time, while the family dynamics and the mental functioning sections could be said to represent more enduring functions, with historical roots (family) and developmental portents (Mental Functioning). This distinction is not unequivocal by any means, but it is the only tentative explanation which suggests itself from the data at hand.

There are two extraneous factors which should be noted as possibly contributing to the area effect. One is the special interest of the interpreter (Henry) in the field of peer interaction, as well as his sharing of the projective tester's "occupational interest"—a particular concern with interpersonal dynamics. Henry's (1956) concept of the shaping of personality by the reciprocal processes of socialization and individuation reflects this double interest. The second extraneous factor which might contribute to the observed area effect pertains to the *S*s of the study. During adolescence, a great deal of psychic energy is bound up in two processes: (*a*) the establishing of relationships with the peer group, and (*b*) the psychic civil wars of emerging sexuality, developing self-concept, and the formalizing of attitudes and values. It is possible that the greater accuracy of TAT interpretation and/or judge decision in these two areas is an artifact of the life stage of these *S*s. A greater emphasis on and intensity of the Peer group and Emotional Adjustment issues may have sharpened the contrasts between individuals for the related statements.

---

[17] Area scores: Areas 1 and 4 show a percentage above chance in I scores of 24.849% and 26.266%, respectively; in D scores, 15.668 and 14.019, respectively. Areas 2 and 3 show corresponding I scores of 13.281 and 14.323, respectively; D scores, of 11.453 and 12.271.

It should be reiterated that the lower efficiency of judgment in the family and mental functioning areas is only a relative matter. Agreement was considerably greater than chance for all four areas. The difference exists, and is striking, but does not invalidate the predictive effectiveness of the TAT in either of the less well-matched areas.[18]

b. Triad. We have assumed the effect of differences between triads to be a "contrast" function. There were some triads in which distinctions between the individuals were relatively great and others in which the distinctions were negligible. This function would also be reflected in any interaction effect to which the triad differences contribute.

c. Trio. Although it is possible to analyze the difference between the three groups of judges as a single effect, this variable is probably confounded by other variables, and for that reason is probably not very meaningful. The different trios are separated by two sets of partitions: The different task sequences and orders assigned to the three trios and the individual personalities which participated in the judgments. Although we have attempted to extricate from this "single" factor the proportions contributed by its definable parts, we have not tried to interpret the trio effect, either singly or in interaction, but have preferred to concentrate on the factors of judge conformity and of presentation effects.

(1) *Judge Conformity.* There are two measures of judge conformity. In the analysis of variance, this factor is represented by the "judge (within trio)" main effect. Of this effect, we can note merely that it is the second largest effect in the combined I and D scores (its differential influence on I and D scores is discussed above). The other measurement of judge conformity is the correlation measure, based on the similarity of decision on matching each individual statement between pairs of judges from each of the three trios. These phi coefficients for judge agreement on performance of identical tasks are surprisingly low.

The psychological implications of the nonuniformity refer principally to the practical clinical use of the TAT. The design of this test actually approximates the clinical situation: the diagnostic test interpretation is presented to a clinician who must—on the basis of that report—gauge the needs and probable reactions of the patient, both as a whole and as a unique individual (the identification task) and as distinguished from all other individuals (the discrimination task). The decisions expected of the clinician are very like those asked of the judges in this study. The results of this study propose that, although expert clinicians are able to make the correct decisions in each case, each one will arrive at the goal by a slightly different route.

(2) *The presentation effects can be considered to be composed of "Sequence" and "Order."* Order refers to the number of judging tasks performed previous to the task in question, and Sequence refers to the nature of the previous tasks.

(*a*) *The Sequence Effect* was not included in the analysis of variance summary (as noted earlier), because of its multiple contamination. The 10 sequence means are based on different numbers of scores, and are confounded by order effects, as well as by the effects of all the "instrument" interactions. We felt that the labor involved in analyzing this effect would only be justified by some evidence of significant influence. Our first (rough) exploration of this factor revealed no significant contribution, and the analysis was not carried further. However, the *F* values obtained for this unfactored effect (using the residual mean square as a denominator) approached significance in both I and D tables, and some discussion of the results is justified.

---

[18] A study by Graham (1947) focused on just the mental functioning area of these same data confirms the predictive efficiency of the TAT in this less efficient area. In this study the correlation between blind IQ estimate from the TAT and Stanford-Binet IQs was .84.

The effect of sequence was among the most powerful in the exploratory study for this research. It appeared that the discriminations made in the immediately preceding task could facilitate or impede the subsequent judgment—as if the concept suggested in considering one criterion either sharpened or clouded the perception of the next. It was for this reason that the exploration of sequence effects was undertaken. In summary,[19] the results of that exploration (from most effective sequence to least) was:

1. Test preceded by Objective and Subjective matching
2. Test preceded by Objective and Projective matching
3. Test preceded by Projective matching
4. Test first in series (not preceded by any other test)
5. Test preceded by Subjective matching
6. Test preceded by Projective and Subjective matching
7. Test preceded by Subjective and Objective matching
8. Test preceded by Projective and Objective matching
9. Test preceded by Objective matching
10. Test preceded by Subjective and Projective matching

The differences in effectiveness of the sequences is in only two cases very large, the first two scoring considerably higher than the remaining eight sequences. Both of these sequences specify the Objective matching as one of the preceding tasks; thus it is possible that the findings are confounded by a tendency of objective matchings to be somewhat less accurate than the other two (since the test performance which measures the sequences' effectiveness is the last test in the series). However, the lack of significance of the instrument effect (test sheet type) prevents us from stating this as a definite conclusion. The only conclusion which can be reached from these data with reasonable confidence is that there is some conflict between any preparatory "sharpening of perception" and what might be called a "fatigue effect," since the effectiveness of "no preceding test" is relatively high in the series.

(*b*) *Order:* The "fatigue effect" mentioned above is confirmed by the scores resulting from different position in the total task. Between the first, second, and third sections of the total matchings assigned, judges show a consistent—though not extraordinarily large—decline in performance. This applies to both I and D scores, and seems to represent a simple function of tiring.

(*c*) *Individual Subject:* The differential effect of particular individual *S*s on the matching performances of judges may be enhanced by contrast factors of the total triads. However, it is more likely to be caused by more personal, subjective (from the standpoint of the judges) considerations, such as the amount of rapport which a judge could establish with a particular *S*, or the inherent interestingness of that *S* to that judge. The fact that the within-triad effect is substantially greater for the I score than for the D scores points up this line of reasoning. If the effect resulted to any considerable extent from degree of contrast, the effect should be more prominent in the task of discrimination.

(*d*) *Instrument:* One negative finding is fully as important as the positive results of any other of the tested variables; this is the nonsignificance of the difference in criterion instrument (test sheet type), and of the instrument-area interaction. It was confidently hoped that the results of this validation study would provide some clue to the place of the TAT in a clinical test battery for personality assessment. Three different types of instrument were included among the criteria against which the TAT was matched, and there was reason to believe that the TAT would show itself to be closer to one than another of these techniques. In stating the evidence acceptable for confirmation of the validity hypothesis, a "standard" was selected from the three criteria for each area of prediction.

---

[19] The actual block scores for the various sequences will be found in Appendix B, among the analysis of variance tabulations.

This selection was made in the express anticipation of notable "area-instrument" interaction. The nonsignificance of these effects presents three alternative interpretations:

(1) There is actually no difference between the reports made from the objective, subjective, and projective tests.

(2) The TAT is so flexible and comprehensive that it taps all levels of all areas equally well.

(3) The interpreters of each of the criterion instruments knew their tests so well that none of them ventured beyond the information legitimately provided by those instruments. If this is so, the criterion information would vary in richness and clarity (as the information sought is nearer or further from the focus of the particular instrument), but need not vary in accuracy; i.e., the statements provided by an instrument in a "less compatible" behavior area might be sparse or vague, but would not be likely to be false statements, because they would not be based on overextension of the instrument in question. Thus, judges might be required to make decisions on the basis of superficial or meager information, or might be faced with a "no information" item, but would probably not be asked to match the TAT with an incorrect statement. Consequently, assuming a consistent level of validity for TAT interpretation, there might be no significant change in error score, but a great deal more uncertainty.

1. The first alternative (equivalence of tests) we can discard, since it is contradicted by all of clinical tradition and experience.

2. The second alternative (flexibility and comprehensiveness of the TAT) is very tempting, and may well be partially true. However, the "one perfect" clinical test idea does not agree with personality theory, with test theory, or with the experience based opinions of testers.

3. The third alternative (limitation of the scope of interpretation)—or at best combination of the second and third—is probably the truest explanation of the unexpected negative finding. We have one index to the plausibility of this explanation. Although the hypothesized "meagerness" and "superficiality" of data are matters of opinion, which it would be difficult to verify, the "no information" items are perfectly definite and verifiable. The distribution of no information items[20] lends credence to our answer. Using the same "standard" criteria for each area (which we defined in the validity hypothesis), we find the "no information" items distributed according to this pattern. The instrument–area combinations chosen as ideal have the smallest number of no information entries. The smallest proportion of "no informations" given for the "subjective (observation and interview)" criterion was in the Family relations area; the smallest proportion of "no information" items in the Emotional Adjustment area is from the "Projective (Rorschach)" criterion; the smallest proportion of "objective" "no information" items is found in the Peer Area. The single contradictory entry in the table is the large number of no information items found for the objective criterion in the mental functioning area. However, 70% of these data failures occurred on Items 14 and 15, which concern creativity and fantasy, and about which the objective data might be expected to have less to say.

CONCLUSIONS

This study represents an attempt to evaluate with maximal precision the effectiveness of a particular projective test in the study of personality. The over-all goal of the research was to arrive at as exact a statement as possible, as to the amount, kind, and degree of accuracy of the information supplied by the Thematic Apperception Test.

The conclusions drawn from the results of this study have to do with two aspects of Thematic Apperception Test validity: (*a*) the accuracy with which predictions can be made from the instrument and (*b*) the use to which those predictions can be put. The actual scores, based on judge decisions, provide the information for evaluation of the TAT; and the way in which the judges made their decisions—using both empirical and inferred evidence of judge performance—offer clues to the way in which TAT reports are actually used clinically.

---

[20] Appendix B.

Conclusions about the Test

1. *The TAT as a valid diagnostic instrument.* In a sense, this was the central question asked by the research. However, not even critics of the test's usefulness seriously doubt that the personality picture it provides has a better than chance agreement with any dependable criterion of the subject's personality. The disproving of the null hypothesis of only-chance agreement is "central" largely because had it not been disproven, the remainder of the findings would have been meaningless. That TAT criterion agreement was so highly significant, in a study where differentiation between subjects was so very minute, makes this finding somewhat more valuable. But it is generally conceded that all thoughtfully designed personality assessments can probably produce a psychological portrait which has some likeness to the model. What we *can* say, by way of original contribution from this single finding, is that our methodological procedures and statistical analysis have apparently translated the clinical meaning of the data successfully into the "language" of experimental method and statistical inference. We say this because the evidence provided by this study about TAT validity agrees with the clinical finding that "it works." Experimental statistical studies of projective test validation have not always been able to demonstrate this conformity.

2. *The TAT is a reliable technique in personality assessment.* As noted in the discussion of reliability, we are only justified in declaring the stability aspect of the concept of reliability, i.e., that the personality revealed by the TAT expresses the constant needs, dispositions, and attitudes of the subject. About the consistency of the test findings from administration to administration or from subject to subject and about the consistency of interpretation of test findings, we can make no unequivocal statement. However, the concepts of stability in time and consistency of findings—though empirically separable—are not independent. If it is demonstrated that a personality test reveals underlying trends in the personality which are not seriously affected by circumstantial changes, or growth, or aging, we can be at least more confident that the test will tend to produce essentially the same information, from administration to administration. This is because the resistance to situational factors is being measured in both cases.

3. *The TAT can predict accurately, in all behavior areas, but is more accurate for some than for others.* We have inferred from the behavior areas of greater and lesser agreement that the TAT tends to reflect more the current conscious and unconscious concerns of the subject than the purely unconscious and deeply repressed sources of these concerns. This conclusion is somewhat confounded by the fact that, for the subjects in this study, emotional adjustment is itself one of the focal concerns. Because there was no significant difference in amount of agreement with the different criteria, it is not possible to assert whether the TAT consistently predicts more accurately at any one particular "level" of personality function.

Conclusions about the Clinical Use of the TAT

The task required of the judges in this research is very similar to the clinical task of utilizing a psychological test report. In planning a therapeutic program, the clinician must decide from the information in the report: (*a*) The nature of the subject's current social milieu (in order to fit in the current fact of therapy); (*b*) The family constellation (in order to select the best possible therapist and the best possible interpersonal approach); (*c*) The intellectual strengths of the subject (to estimate the amount of help the therapist can expect and to decide on a realistic aspiration level); and (*d*) The subject's emotional adjustment (for a detailed picture of the foci of the subject's difficulty, his personal strengths and weaknesses, and the probable prognosis for, and course of, treatment). Although the goals were not the same, and the judges may have been less motivated to make the correct decision than in the clinical setting, the actual judgments made were substantially those which would be made by the same judge clinically. Thus, there is some justification for inferring from the characteristics of the performances of these expert judges in the task at hand to the performances of these (or other similarly trained) clinicians, in clinical utilization of the TAT report. There were several rather striking trends in the judges' responses, which lead us to conclude:

1. *There is no single correct way of employing the TAT interpretation.* There was little item agreement between judges, but each judge made enough "correct" decisions to yield a highly significant agreement figure. Judges may arrive at essentially the same interpretive implications of the test report, by quite different routes; or judges may differ individually in their ability to utilize TAT predictions in different areas (supported by the finding of significant trio-area interaction) or for different subjects (possibly supported by the significant trio-triad interaction, although the triad effect includes a differential "contrast" factor as well). But whatever the route used or the implications noted, all judges arrived at a substantially accurate understanding of each subject, despite great interjudge variation.

2. *Judges (clinicians) appear to make more consistent use of the projective (TAT) report when the decisions they must make are definitive and detailed* then when they are relatively simple, i.e., because there is overlap in the behavioral repertoires of individuals, it is easy to make the mistake of including in a subject's potential responses those which are more characteristic of someone else. It is more difficult (and more inaccurate) to exclude from a subject's potential responses those he in fact displays. Conversely, it is more difficult to make the first type of judgment correctly than the second. Although there were, as expected, more mistakes of the first type (D errors), the judges varied much more widely in their ability to make the second type of judgment (I judgment) accurately. We have concluded that the "set" necessary for making fine distinctions (concentration, deliberation, and impersonal, detached judgment) is more conducive to consistently accurate diagnosis than the more spontaneous, personally colored approach used in making less tedious comparisons.

3. *The judges' perceptions were sharpened by consideration of a variety of personality data, but there is a point of diminishing returns.* Although a certain amount of facilitation in judgment is produced by examining the subject from several angles, there is also a "fatigue" effect produced by a superfluity of information, and this effect becomes dominant when the clinician considers many subjects.

4. *Clinicians have their particular "specialties."* The accuracy of matching differed, with each judge, from individual to individual. There was, at the same time, no clear agreement between judges as to which individuals were judged more accurately. The exact cause of the divergence is not clear, but the crucial variable is apparently strictly personal. That is, it is explained by a difference in the way a judge perceives certain subjects or groups of subjects. It may be that the judge is especially expert at recognizing all hysterics; another judge matches boys better than girls; a third simply finds certain individuals more interesting than others, and is thus more involved with those individuals. In any case, each judge fluctuated in the ability to judge different individuals, and no two judges fluctuated in exactly the same direction.

A projected subsequent study, based on judges' responses to follow-up questions about attitude, work habits, etc., in the present research, may help to refine the conclusions stated somewhat tenuously in this section on the clinical use of the TAT.

REFERENCES

American Psychological Association. Committee on Psychological Tests. *Technical recommendations for psychological tests and diagnostic techniques.* Washington, D. C.: American Psychological Association, 1954.

Brown, A., Morrison, J., & Couch, Gertrude. Influence of affectional family relationships on character development. *J. abnorm. soc. Psychol.,* 1947, **42**, 422–428.

Combs, A. W. The validity and reliability of interpretations from autobiography and the Thematic Apperception Test. *J. clin. Psychol.,* 1947, **3**, 257–264.

Cox, Beverly, & Sargent, Helen. TAT responses of emotionally disturbed and emotionally stable children: Clinical judgment vs. normative data. *J. proj. Tech.,* 1950, **14**, 61–74.

Cronbach, L. J. A validation design for qualitative studies of personality. *J. consult. Psychol.,* 1948, **12**, 365–374.

Davis, Hilda A. Moral ideology and attitudes of Indian children. Unpublished doctoral dissertation, Univer. of Chicago, 1953.

Graham, Elaine. A comparison between estimates of intelligence from projective and standard tests. Unpublished master's thesis, Univer. of Chicago, 1947.

Havighurst, R. J., & Neugarten, Bernice. *American Indian and white children.* Chicago: Univer. Chicago Press, 1955.

Havighurst, R. J., & Taba, Hilda. *Adolescent character and personality.* New York: Wiley, 1949.

Henry, W. E. *The analysis of fantasy.* New York: Wiley, 1956.

Kogan, L. S., & Hunt, J. McV. Problems of multi-judge reliability. *J. clin. Psychol.,* 1950, **6**, 16–19.

Krugman, Judith I. A clinical validation of the Rorschach with problem children. *Rorschach Res. Exch.,* 1942, **6**, 61–70.

Norton, Dee Wiswell. An empirical investigation of the effects of non-normality and heterogeneity upon the F test of analysis of variance. *Dissertation Abstr.,* 1952, **12**, 713–714.

Rosensweig, S. Apperceptive norms for the Thematic Apperception Test: II. The problem of norms in projective methods. *J. Pers.,* 1949, **17**, 475–482.

(Early publication received June 25, 1959)

APPENDIX A

FORMS AND TESTING MATERIALS

TABLE 3

PRESENTATION ORDER OF JUDGING TESTS

TRIO 1 (Judges 1, 4, 7)			TRIO 2 (Judges 2, 5, 8)			TRIO 3 (Judges 3, 6, 9)		
Triad	Subject	Test Sheets	Triad	Subject	Test Sheets	Triad	Subject	Test Sheets
I	A	5-8-2	VII	C	1-4-7	XII	B	8-2-5
II	C	9-3-6	VIII	B	8-2-5	XI	C	4-7-1
III	B	1-4-7	VI	A	6-9-3	X	A	3-6-9
IV	B	8-2-5	IV	A	8-5-2	IX	A	2-5-8
V	A	3-6-9	V	B	4-7-1	VIII	A	4-1-7
VI	C	4-7-1	IX	C	1-4-7	VII	B	6-9-3
VII	A	2-5-8	I	B	8-2-5	VI	B	5-8-2
VIII	C	6-9-3	XI	B	1-4-7	V	C	7-4-1
IX	B	7-1-4	III	A	8-5-2	IV	C	9-3-6
X	B	5-8-2	X	C	1-4-7	III	C	8-2-5
XI	A	9-3-6	II	B	5-8-2	II	A	4-7-1
XII	C	1-4-7	XII	A	9-3-6	I	C	3-6-9
I	B	7-1-4	VII	A	9-3-6	XII	A	6-9-3
II	B	2-5-8	VIII	C	1-7-4	XI	B	2-5-8
III	C	6-9-3	VI	C	1-4-7	X	C	7-1-4
IV	A	1-4-7	IV	C	4-1-7	IX	B	3-6-9
V	C	5-8-2	V	A	8-5-2	VIII	B	8-2-5
VI	B	9-3-6	IX	A	3-9-6	VII	C	1-4-7
VII	B	4-7-1	I	C	6-3-9	VI	A	9-3-6
VIII	A	8-2-5	XI	A	6-9-3	V	B	5-8-2
IX	C	3-6-9	III	B	4-7-1	IV	B	4-7-1
X	A	7-1-4	X	B	9-3-6	III	A	6-9-3
XI	C	2-5-8	II	A	3-6-9	II	C	2-5-8
XII	B	6-9-3	XII	C	5-8-2	I	A	7-1-4
I	C	3-6-9	VII	B	5-8-2	XII	C	7-1-4
II	A	4-7-1	VIII	A	6-3-9	XI	A	6-9-3
III	A	8-2-5	VI	B	8-2-5	X	B	2-5-8
IV	C	6-9-3	IV	B	3-6-9	IX	C	4-7-1
V	B	7-1-4	V	C	3-9-6	VIII	C	3-6-9
VI	A	2-5-8	IX	B	5-2-8	VII	A	8-2-5
VII	C	9-3-6	I	C	1-4-7	VI	C	1-4-7
VIII	B	1-4-7	XI	C	8-2-5	V	A	9-3-6
IX	A	5-8-2	III	C	6-9-3	IV	A	5-8-2
X	C	3-6-9	X	A	2-5-8	III	B	7-1-4
XI	B	4-7-1	II	C	7-1-4	II	B	6-9-3
XII	A	8-2-5	XII	B	4-7-1	I	B	2-5-8

THE JUDGING TASK

A sample judging task is presented in the immediately following pages. The *S* is one of those actually used in this study. Materials are given in exactly the form and the order in which judges received them, and samples of all materials are included:

1. Instructions to Judges ("Information for Judges")
2. Outline of Personality Areas

3. TAT Summary
4. Criterion Statements in Peer, Family, Mental Functioning, and Emotional Adjustment Areas, facing TAT Summaries of these Areas
5. Sample Judging Form

SAMPLE SCORING FORM

	TRIAD____			TRIAD____			TRIAD____			TRIAD____			
	Subject____			Subject____			Subject____			Subject____			
	Test Sheet			Test Sheet			Test Sheet			Test Sheet			
1													1
2													2
3													3
4													4
5													5
6													6
7													7
8													8
9													9
10													10
11													11
12													12
13													13
14													14
15													15
16													16
17													17
18													18
19													19
20													20
21													21
22													22
23													23
24													24
25													25
26													26
27													27

Fig. 1 Judging Sheet.

INFORMATION FOR JUDGES

Each of the enclosed sets of papers contains:

1. An outline of the personality areas covered in this study.
2. A complete analysis of a TAT protocol, based on this outline.
3. A set of statements, also based on the personality outline, but derived from sources other than the TAT. The source is indicated, in each case.

What we are asking of you is, broadly speaking, a matching operation, involving a TAT interpretation compared against a set of statements, both covering the same areas of functioning. Some of the statements on each statement sheet will describe the individual whose TAT interpretation is presented, and some will not. You are asked to decide which of the statements do describe the TAT *S*. Each statement sheet is made up of a composite of statements about three individuals, one of whom is the *S* of the TAT analysis. Items for the statement sheet have been assigned randomly and independently from the separate statement sheets of these three individuals. Thus, the proportion of statements referring to each individual, as well as the order and position of items describing a particular *S*, varies from sheet to sheet. It is possible, for example, for an entire statement sheet to contain only items referring to the TAT *S*; similarly, it is possible for a particular statement sheet to include no items at all which describe the subject in question—or any proportion between these extremes.

Instructions

1. Before beginning the actual judging, it will probably be useful to study the personality outline, in order to familiarize yourself with the areas of comparison with which you will be dealing.

2. Before starting a particular set of scorings, we ask you to read carefully the complete TAT summary that precedes it, so as to acquaint yourself with the total personality of the individual, against which the statements are to be matched.

3. Rating:

a. In addition to the complete analysis, the TAT summary is also reproduced, area by area, to correspond to a cluster of statements facing it. In each area, we ask you to read over the part of the TAT summary which faces the statements which you are rating.

b. Then read the first statement and judge whether, in your opinion, it describes the *S* of the TAT. If you believe it does, mark a "+" in the appropriate space on the Rating Sheet. If you believe it does not, mark a "0" in the space. If you are unsure, score the item "?." (*Note:* all statements which are listed as "no information" may be scored "?.")

c. Make the judging of each statement as independent as possible of decisions you have reached on previous statements.

d. Keep in mind, in scoring these statements, the fact that the population used in the study is equated for sex, age, education, and—roughly—for socioeconomic factors (all the *S*s were selected from the same school). Differences may be slight and subtle and similarities frequent.

e. Keep in mind also that the items scored "?" are statistically less valuable than "+" or "0" items. Do not use a question mark if you can possibly reach a decision.

4. Check each set of statements to make sure you have scored all of them.

OUTLINE OF PERSONALITY AREAS

I. *Social Interaction*

A. *Peers*

1. Degree of participation (integral member, isolate, active participant, number of mutual choices, social success, amount of acceptance, etc.)

2. Nature of participation (leader, follower, conforming, rebelling, dependable, enthusiastic, spontaneous, etc.)

3. Quality of relation to peers (group warm/cool towards him, stimulus value high/low, warm/cool feeling toward others, trusted/not-trusted, attacking, nonattacking, self-justification needs, etc.)

B. *Family*

4. Feeling toward mother and/or personality of mother
5. Feeling toward father and/or personality of father
6. General family atmosphere (emotional attitudes, kind and degree of regularity, etc.)
7. Sibling relationships
8. Relationship between mother and father

C. 9. Social relation to the Community

II. *Characteristics of the Self*

A. *Mental Functioning*

10. *IQ* (potential functioning)/present level of functioning (school work, efficiency, etc.)

11. Mental approach and organization (judgment of level of organization and qualitative elaborations over/under-generalize, over/under-relate, compulsive trend, premature closure or attention to wholes only, etc.)

12. Motivational factors (ambition, energy investment in work, positive interest in achievement, accomplishment, conscientiousness, over/under-scrupulousness, etc.)

B. *Imagination*

13. Quality of imagination and creativity (high/low/mediocre creativity, originality, amount of restraint/freedom, etc.)

14. Fantasy as an escape (some, none, limited, much, etc., use of fantasy as escape.)

15. Nature of fantasy—in general (stereotyped, self-glorifying, omnipotence, etc.)

C. *Patterns of Emotional Adjustment*

16. Basic emotional attitude (active/passive)

17. Impulse life acceptance (accept/reject, important/unimportant, etc.)

18. Anxiety

19. *a.* Presence and degree, or absence, and form taken
    b. Source or primary locus

D. *Mechanisms of Defense (Control System)*

20. Conscious control/inner control (behavior directed primarily by conscious consideration/inner needs)

21. Outer control (amount and ease of conformity with outside demands)

22. Special mechanisms (comforting devices: achievement, fantasy, sensuousness, etc.; other ego defenses: isolation, reaction formation, etc.)

E. *Emotional Reactivity*

23. Inner amount of introversion (extent of inner preoccupation, and possibly more specific focus, reason, or nature of inner preoccupation)

24. Outer-extroversive tendencies (amount of interest in the outside world, things/people, action/thinking, etc.)

25. Potentiality (what is likely to happen to the balance of 23/24)

F. *Sexual Adjustment*[a]

26. Anxiety and general adjustment (anxiety, hesitancy, fear, etc., amount and kind, excepting comments more appropriate to 27, below)

27. Role (aggressiveness/passivity, amount of interest, participation: social/phallic, etc.)

TAT SUMMARY

Triad IV

Individual A

I. *Social Interaction*

A. *Peer relationship.* His peer relations should not be good. He is still too concerned with parental problems and how to conform to work out ties with peers. His control is quite gross and his techniques are equally inept.

Overt pattern should be one of some conflict with others. This arises for several reasons:

1. His strong ego and strong inner life
2. His resistance to parental demands
3. His guilt over his resistance

S is a boy both mature and immature—of strong undisciplined impulsivity that is suppressed by his super-ego-demands of parental conformity. In his overt behavior his conflict will be of the willful, resistant, and demanding sort.

B. *Family Relationships*

Intrafamily dynamics. Mother relationship is a difficult one. The tie is largely a fixation wherein his own spontaneity is viewed as rebellion against the mother and hence arouses guilt. The mother seems rather subtly demanding and oriented more toward achievement demands than to repressive impulse control.

---

[a] In analyzing the judge decisions, Items 26 and 27—Sexual Adjustment—were considered to be part of the "Peer Interaction" Section and were so counted.

The father tie is a very difficult one in that he is a man of firm and uncompromising demands which *S* tries to obey. The boy views him as demanding and firm and also as opposed to *S*'s affectionate interest in the mother. It is the father who keeps him away from the mother, he feels.

His high narcissism and infantile emotionality reflects a fixated primary tie to the mother marked not by overaffectionate protection, but by its lack. I presume the tie to have been rather routine and unloving in the early years.

The continued longing for affection, plus the high awareness of control, plus the reprisal for self-affront (3BM-18BM) suggest the home atmosphere has been one oriented by strong-willed parents more concerned with their own interests (i.e., mobility) than with the affection of their children.

II. *Physical Characteristics*

A. *Energy Output.* *S* has a rather strong emotional energy output, the expression of which is made with rather gross manipulatory, large-muscle movements.

B. *Growth Pattern.* The TAT suggests that he is post-pubescent and actively oriented heterosexually.

III. *Characteristics of the Self*

A. *Mental Functioning.* *S*'s intellectual level is superior (120-125 IQ). His intellectual efficiency, however, is at present greatly reduced. His emotional preoccupations are no doubt responsible partly for this. He shows a lack of ability to discipline his concepts and a low use of whole inclusive concepts though he can be forced into them. He prefers details and is a keen observer of them (Pic. 18).

School adjustment could be very good under certain circumstances where a firm but obviously affectionate teacher could encourage his rich inner life and have him discipline himself. His low efficiency, however, is not conducive to good school work in the traditional sense, and he is not really interested in intellectual matters. His interests lie much more in the manipulation area (Pic. 2).

B. *Imagination.* *S*'s imagination is a good quality but hardly functional because of his poor discipline and organization and because of his guilt over any form of spontaneity or impulsivity.

He has considerable potential here that is hardly being used.

Active fantasy life that takes up some of his energy but that is not completely effective as control.

C. *Patterns of Emotional Adjustment.* Basic emotional attitude is slightly aggressive. He maintains a strong ego that is hampered by guilt but that is still struggling actively.

Impulse life is rich and accepted. It is confused by his guilt over nonconformity, but this is not impulse-repression.

S has many signs of anxiety and feelings of insecurity, particularly in his peer and in his parental ties. He describes people separately and does not combine them into cooperative congenial groups. His descriptions of people are also generally of restraint and conformity, of worry, of disappointment.

System of control is heavily weighted to the introversive. His conscious control is high (not constricted) and his outer control is only partially effective. Inner control is used but his resources here are so rich that his ego-demands continually overburden his outer control.

Interpersonal relations are his bad point. His view of people as either not loving enough or as demanding makes it hard to establish contact. He is so preoccupied that he is also not making much effort at this area (low c in Ror?).

S has sexual interests that are developing. His infantile attachment has not been impulse-repressing in light of the achieving-demanding (rather than highly repressive) mother.

IV. *Summary*

This is a boy currently preoccupied with the following themes:

1. Longing for protection and love from mother
2. Resistance to accepting demands
3. Guilt over his own resistance

He is a boy of superior intelligence and high creativity who is so restrained and apprehensive over his "self" that he has no energy left for working out interpersonal ties. The dominant causative factors suggested are:

1. Early lack of maternal love
2. Too rigid achievement demands
3. Father attitude too firm and demanding
4. Early lack of emotional contact with father

Sample Matching Form

The form for matching presented to each judge was stapled in such a manner that the TAT summary for that area appeared on the left and the corresponding statements to be matched appeared on the right.

Triad IV

Individual A

A. *Peer Relationship*

His peer relations should not be good. He is still too concerned with parental problems and how to conform to work out ties with peers. His control is quite gross and his techniques are equally inept.

Overt pattern should be one of some conflict with others. This arises for several reasons:

1. His strong ego and strong inner life
2. His resistance to parental demands
3. His guilt over his resistance

S is a boy both mature and immature—of strong undisciplined impulsivity that is suppressed by his super-ego-demands of parental conformity. In his overt behavior his conflict will be of the willful resistant and demanding sort.

Test Form 2 (Sociometric)

Peer Area:

1. "No evidence of spontaneous participation."

   " . . . a high score on leadership when he apparently does not participate very actively or effectively in sports. . . . "

   " . . . very actively participating in the group. . . . "

2. "At no time did the *S* have a positive role in the group."

   ". . . appeared almost in a scapegoat role. . . . "

   " . . . has presented two fronts, first extreme hostility and later indifference or withdrawal."

3. " . . . very genuine warmth and trust in him expressed by the group."

   " . . . prestige in the group is so clearly defined that one has a strong impression that he has been enjoying a top status position in a group throughout his career as a participant in it."

   " . . . appears to be a person who can do no wrong."

B. *Family Relationships*

Intrafamily dynamics.

Mother relationship is a difficult one. The tie is largely a fixation wherein his own spontaneity is viewed as rebellion against the mother and hence arouses guilt. The mother seems rather subtly demanding and oriented more toward achievement demands than to repressive impulse control.

The father tie is a very difficult one in that he is a man of firm and uncompromisive demands which *S* tries to obey. The boy views him as demanding and firm and also as opposed to *S*'s affectionate interest in the mother. It is the father who keeps him away from the mother, he feels.

His high narcissism and infantile emotionality reflects a fixated primary tie to the mother marked not by overaffectionate protection, but by its lack. I presume the tie to have been rather routine and unloving in the early years.

The continued longing for affection, plus the high awareness of control, plus the reprisal for self-affront (3BM-18BM) suggests that the home atmosphere has been one oriented by strong-willed parents more concerned with their own interests (mobility) than with the affection of their children.

Test Form 8 (Observation and Interview)

Family Dynamics

4. Obs.: "The mother serves apparently as a buffer between the father and the son. However, there is a strong indication of affection by both. Mother . . . works in the family's shop, painting and doing other forms of manual labor. In addition does some of the office work . . . Mother states views as to man's dominant place in the home. The mother is small and pretty and showed a great deal of cooperation to the interviewers."

   Int.: "She talks with great intensity clasping her hands before her and sitting on the edge of her chair. She seemed very much moved at times by things she was telling me . . . a very pleasant, charming person who undoubtedly is very affectionate with her children . . . "

5. Obs.: "The father has an autocratic attitude . . . The father demands a good deal of (*S*) in his shop work and in general seems prone to ignore good work and criticize failure . . . (*S*) is especially sensitive to his father's disapproval . . . There is frequent reference to the

father and son siding together against the mother and sister. This takes the form of critical teasing."

Int.: " . . . a tall heavily built young man with a face rather like (*S*'s). He was very cordial and friendly . . . By his conversation and manners he seemed an affable sort of fellow, not overbearing or domineering in any form . . . Mr. ________ never paid much attention to the children when they were little . . . Mother says that (*S*) has a profound respect and affection for his father . . . The impression I gained is that his father treated him as a complete equal in many instances, but that when there was a difference of opinion (*S*'s) opinions were given no weight and he was secretly pleased whenever he was right. . . . (*S*) spoke with admiration of his father and his remarkable memory at the shop although in the same incident he mentioned how his father would give his commands sharply and shortly to him."

T.: Reporting an account by (*S*'s) sister of father's reaction to (*S's*) running away. " . . . he ran away from home about a month ago . . . her dad was so mad he said not to even look for him this time. I guess they have a terrible time getting along . . . she says her mother tries to make peace between them, but she can't do a thing."

M.: " . . . his father was so angry with him most of the time. He just wouldn't stand for (*S*'s) acting that way and they just couldn't get along . . . he says (*S*'s) doesn't know how to handle money so he shouldn't be allowed to have any . . . His father will say to him, rather sharply, 'Now just what do you think you're doing it that way for?' and (*S*) will always have a good reason. His father will mutter under his breath to anyone else, 'He's a better mechanic than most any grown man!' . . . his father won't stand for any nonsense from him."

6. Obs.: "(*S*) goes to the store. Carries out ashes. Takes care of the garbage. Helps father . . . Probably held near home by parents, but somewhat free there."

Int.: "The family lives . . . in a relatively new white frame house . . . the interior is very attractively painted and furnished . . . almost entirely due to the efforts of (mother) and her husband. (Parents) very interested in children and their education. Children have chores and are expected to do them."

T.: " . . . the family are above average in wanting (*S*) to do things. They attend all the basketball games. They show an interest in what the children are really doing and talking about, too."

7. Obs.: "Has one sib—a sister, 14. Each have their own friends, do not do things together. They fight some, but stick up for one another."

T.: "He teased his older sister, but a genuine affection has always existed between them despite bickerings . . . no deep jealousy or animosity between them."

8. Obs.: "Mrs.__________ was always overshadowed by her husband and brothers."

9. Sub: "I like this place. If I lived in the country there wouldn't be anything to do. If I lived in a city like Chicago I wouldn't have a yard and I'd live in an apartment.

TAT Summary

C. *Patterns of Emotional Adjustment*

Basic emotional attitude is slightly aggressive. He maintains a strong ego that is hampered by guilt but that is still struggling actively.

Impulse life is rich and accepted. It is confused by his guilt over nonconformity, but this is not impulse repression.

S has many signs of anxiety and feelings of insecurity, particularly in his peer and in his parental ties. He describes people separately and does not combine them into cooperative congenial groups. His descriptions of people are also generally of restraint and conformity, of worry, of disappointment.

System of control is heavily weighted to the introversive. His conscious control is high (not constricted) and his outer control only partially affective. Inner control is used but his resources here are so rich that his ego-demands continually overburden his outer control.

Interpersonal relations are his bad point. His view of people as either not loving enough or as demanding makes it hard to establish contact. He is so preoccupied that he is also not making much effort in this area (low C in Ror?).

S has sexual interests that are developing. The fact that his infantile attachment has not been impulse-repressing is interesting in light of the achieving-demanding (rather than highly repressive) mother.

Test Form 5 (Rorschach)

Patterns of Emotional Adjustment:

16. Basic emotional attitude shows signs of constraint. There is . . . a reluctance to admit the existence of emotional problems . . .

Underlying submissive tendency possibly which is in conflict with what he, as a boy, is expected to do.

17. Restriction of childlike, concept-dominated impulse life in terms of constrictive control functions, both of which exist side by side, as it were, and provide constant element of conflict.

18. While there are evidences of anxiety, chiefly in terms of negativism, it is an anxiety which is more of the situational-reactive kind rather than some deep-seated disturbance that has taken on systematic defense patterns.

19. As the source of anxiety, we might postulate guilt over resistance to paternal control, ambivalent feelings towards father, and possibly a component of covert hostility, aside from concern about heterosexual relations.

20. Chief motive direction, inner.

21. (*S* tends to) do right, conform, obey rules, be conscientious, the proverbial "try and try again" attitude of rigid and righteous perseverance.

22. There is a good possibility of *S*'s seeking (i.e., propelled) sensuous aesthetic pleasure through daydreaming, oral satisfactions, and through gaining love and affection from females who have sufficient maturity to become a mother substitute.

23. There is an underlying introversive trend.

24. At the present time, he probably does not allow himself really to become involved in interpersonal relations, due to preoccupation with inner emotional problems and as a protection against infringement of ego integrity.

In terms of his childlike emotional dependency, which is denied in most instances, I would not think that S is capable of close interpersonal ties at this time.

I would expect him to form fleeting associations to get emotional support; i.e., he'll associate with different pals to do things, rather than choosing people to be with.

25. No information.

26. There seems to be considerable anxiety with regard to sexual impulses . . .

27. In terms of progression into social sex role, he is decidedly in a period of transition, somewhat handicapped by his relative emotional-social immaturity.

EXAMPLE OF "HIGH-CONTRAST" MATCHING

The following set of data is an instance of high constrast between *S*s. The TAT summary and criterion statement obviously refer to different *S*s.

As before, the TAT summary is given to the left, the incorrect and correct criterion statements are listed to the right, with the correct statement immediately below the incorrect one:

Triad I—Individual A

Family Dynamics

The family picture is a rather detached and impersonal one. The dominant feature should include a mother to whom there may or may not be some hostility, but who is seen somewhat as an impersonal nuisance. The mother is clearly present and accepted only as a controlling person. It is as though the mother had very low emotional contact with him, but did assert herself so that he cannot afford to ignore her or be hostile to her. The father relation seems equally or even more distant, yet at this distance, seen as a desirable object. He has almost a devotion to this distant father image. The total family picture does not stress warmth or unity of interpersonal relations. There is the distinct probability of direct sibling hostility.

Triad I—Test Form 7
(Subjective—Interview & Observation) (incorrect)

Family Dynamics

6. *Obs:* The house is quite an old house, but attractively furnished—a good many books in the living room, magazines, etc. . . . one gets the impression of this family as very secure, affectionate home atmosphere, in which the parents are secure and permit the children to become so. Apparently they take real pride in the children's behavior in school. Both Mr. and Mrs. __________ are very easy to talk to and talk freely and naturally. We did not confine our conversation to discussion of the children, but talked about national affairs, books, etc.

Triad I—Test Form 8
(Subjective—Interview and Observation) (correct)

Family Dynamics

6. *Obs:* Family lives in a small one story house which looks well kept up from the outside. Inside, it is extremely disorderly. All of the young children were very dirty, both clothing and faces, and they didn't look as if much attention had been given them. Mother revealed that the kids tear around the house a good deal and have broken the windows many times . . . The babies were always dirty, with runny noses and wet pants.

Mother: I don't let him keep all that money (earned on a job). I usually have him give me the $2 and put it away; he gets to keep the 50 cents . . . He eats pecky, doesn't like vegetables or nothing raw. ("Does he get a piece of fruit every day?") I guess not . . . He drinks (milk) when we have it. He'll drink a quart at a time if you let him.

EXAMPLE OF "LOW-CONTRAST" MATCHING

The Judges were asked to match items taken from one of the criterion tests with relevant portions of a given TAT summary. In some instances, two or more individuals from the same triad showed much similarity in personality structure. The criterion statements of such individuals are necessarily similar also. In those cases of low contrast, even where the criterion and the TAT do *not* refer to the same *S*, the accidental similarity between the two statements might lead the judges to believe that statements had a common referent. The judges' response would probably have been to score the items incorrectly as referring to the same *S*.

In the instance of low contrast cited below, the TAT summary is given to the left, the criterion statements to the right—first the incorrect statement, and immediately below it, the correct one:

Triad II—Individual A

Peer Group Relationships

S's relationship with peers is extremely poor—poor in the sense that he has few positive ties to them and in general he is received negatively by them. This is an instance where the social stereotypes will probably operate strongly against him to give him a reputation of rejection. Actually he has very little hostility against his peers, but his feelings of self-martyrdom, his strong introspection, and his rather sensuous and full inner vitality will be received with suspicion by the group. I should expect that his sociometric scores would be heavily on the rejection side. While he is a good object for scapegoating, I doubt if he received much of that kind of attack at present.

Triad II—Test Form 3
(Objective-Sociometric Test Analysis) (incorrect)

3. " . . . occasional flare-ups of temper . . . "

" . . . is regarded as a weak, distasteful person, not to be trusted."

Triad II—Test Form 2
(Objective-Sociometric Test Analysis) (correct)

3. "The group shuts him out but does not seem to punish."

"His classmates respond to him negatively apparently not for what he does or says but because of some 'queerness' which they find distasteful."

" . . . does not seem to be the object of hostility or scapegoating, but rather of undefined distaste or repugnance."

VALIDITY OF CRITERION INSTRUMENTS

The degree to which criterion instruments "measure what they purport to measure" limits the usefulness of the statements produced by these instruments in testing the validity of the TAT predictions.

The *projective* instrument used (Rorschach) is a well-validated test, in most of the senses in which "validity" is used psychologically. Its dependability in clinical diagnosis, as well as its differentiating power in "normal" *S*s, is too well known and the validation studies too numerous to warrant the listing of such findings here.

The subjective instruments (observations and interviews) are directly extrapolated samples of the behavior level they serve. As such, they do not "measure" or "stand for" the behavior they describe; they *are* that behavior level. A "validation" of the observations and interviews would be meaningless for the purposes of this study.

The same reasoning applies, in the present research, to certain of the *objective* criterion instruments. The Sociogram and the Guess Who Test are felt to yield objective information in the Peer Area. These tests tap directly the stimulus value and group interaction of each *S* with his peers. Mental functioning, in the objective data, is represented by well standardized tests of intelligence and achievement, by school grades, and by teachers' observations (the latter two, like the Sociogram and Guess Who, expressing directly the content with which they deal). However, two "objective" instruments—the Family Relations Questionnaire (yielding data in the Family area) and the Emotional Response test (producing Emotional Adjustment statements)—require comment:

1. *The Emotional Response Test.*[a] This ". . . is a series of questions asked by the tester to elicit from the *S*'s statements concerning activities and persons who produced in them the emotions of happiness, sadness, anger, fear, and shame" (Davis, 1953, p. 14). The information asked of the *S* is introspective,

---

[a] This test was developed for a comparative study of moral ideology and emotional attitudes in three societies (5). Details of administration, scoring, categorization, and reliability testing will be found in this study (pp. 14-23, 623-651); the nonqualified "validational" material dealing with the relationship of ERT responses to observed behavior can also be found in the section of that study which treats results (pp. 350-380, and summarized on pp. 602-612).

and requires the *S* to report feelings rather than exhibit behavior. Such material is subject to more distortion than the content of, e.g., the Sociogram or the Guess Who Test, and we cannot accept it as automatically as a literal statement of the *S*'s feeling reactions.

Measurement of the reliability of the test (defined as consistency of response categories in two matched groups) gave correlation coefficients of .59–.66(+.03) for "Anger," of .63–.86(+.01), for "Sadness," and of .93–.95(+.01) for "Happiness." No formal validation of the ERT has been reported, but its general agreement with the results of other tests used in the same research and its usefulness in elaborating and integrating the psychic hypotheses to which the research was addressed provide sufficient justification for its inclusion in this study. Furthermore, the interpretive procedure and the reliability standardization of the ERT were performed on an age group which is approximately the same as the *S*s of the present study, which should provide maximum relevance for its application to this adolescent population.

2. *The Family Relations Questionnaire.* The validity of the FRQ as a diagnostic instrument has been tested. The responses to the Questionnaire were interpreted to yield information about specific traits of the *S* and about more general personality trends. These interpretations were compared with independent estimates of the traits derived from a battery of other tests (the "Guess Who Test," the "Portrait Guess Who," the "Check List," and the "Character Sketch") and with broader personality descriptions derived from clinical ratings and from more global personality tests (e.g., the California Personality Test), with an average trait correlation of .67 (range .50–.79). These measures of agreement were not ideally high, but are consistent. Since, in the present study, we are using the FR responses without interpretation as a simple report of family interaction, these correlations obtained from more exacting use of the test are ample ground for its inclusion as one of our criterion instruments.

APPENDIX B

TABULAR AND GRAPHIC PRESENTATION OF RESULTS OF THE STUDY

TABLE 4

Analysis of Variance Block Scores

I SCORES

Area	Scores[a]	N	X[b]
1	24154	972	24.849
2	12910	972	13.281
3	13922	972	14.323
4	25531	972	26.266
	76517		
• Standard error (area) = 1.117			
Trio			
1	30460	1296	23.503
2	21551	1296	16.628
3	24506	1296	18.908
	76517		
• Standard error (trio) = .978			
Triad			
1	4642	324	14.327
2	6443	324	19.885
3	5469	324	16.879
4	7834	324	24.179
5	7984	324	24.641
6	5514	324	17.018
7	3386	324	10.450
8	6009	324	18.546
9	9008	324	27.862
10	7585	324	23.410
11	8172	324	25.222
12	4471	324	13.799
	76517		
• Standard error (triad) = 1.944			
Instrument			
1	26532	1296	20.472
2	24078	1296	18.578
3	25907	1296	19.989
	76517		
• Standard error (instrument) = .963			
Order			
1	28507	1296	21.966
2	25102	1296	19.368
3	22908	1296	17.675
	76517		
• Standard error (order) = .980			
Sequence			
0	26147	1296	20.175
1	7349	420	17.497
2	9696	456	21.263
3	8112	420	19.314
4	8957	384	23.325
5	790	36	21.944
6	7835	396	19.785
7	1134	60	18.900
8	5986	372	16.091
9	511	48	10.645

0 Test first in series (for the individual subject)
1 Test preceded by Projective and Objective matching
2 Test preceded by Projective matching
3 Test preceded by Projective and Subjective matching
4 Test preceded by Objective and Subjective matching
5 Test preceded by Objective and Projective matching
6 Test preceded by Subjective matching
7 Test preceded by Subjective and Objective matching
8 Test preceded by Objective matching
9 Test preceded by Subjective and Projective matching

Subject	Scores[a]	N	X[b]
1	2252	108	20.851
2	861	108	7.972
3	1529	108	14.157
4	1669	108	15.453
5	2297	108	21.268
6	2477	108	22.935
7	1798	108	16.648
8	1066	108	9.870
9	2605	108	24.120
10	2870	108	26.574
11	1861	108	17.231
12	3103	108	28.731
13	2706	108	25.055
14	2054	108	19.018
15	3224	108	29.851
16	962	108	8.907
17	3084	108	28.555
18	1468	108	13.592
19	1220	108	11.296
20	557	108	5.157
21	1609	108	14.898
22	1909	108	17.675
23	2199	108	20.361
24	1901	108	18.601
25	2490	108	23.055
26	3637	108	33.675
27	3881	108	26.675
28	3009	108	27.861
29	1853	108	17.157
30	2723	108	25.212
31	2384	108	22.074
32	1609	108	14.898
33	4179	108	38.694
34	1520	108	14.074
35	2107	108	19.509
36	844	108	7.814

• Standard error (subject) = 3.335

---

[a] For definition of "score" see text, pages 29, 30.

[b] Since Chance = 0, the Means Column (X) represents the percentage of correct judgments in each category greater than chance.

• Formula for standard error $\left(\sqrt{\frac{SS_T - SS_Q}{df_T - df_Q}}\left(\frac{1}{NQ}\right)\right)$ assumes equal variances in all levels of the factor.

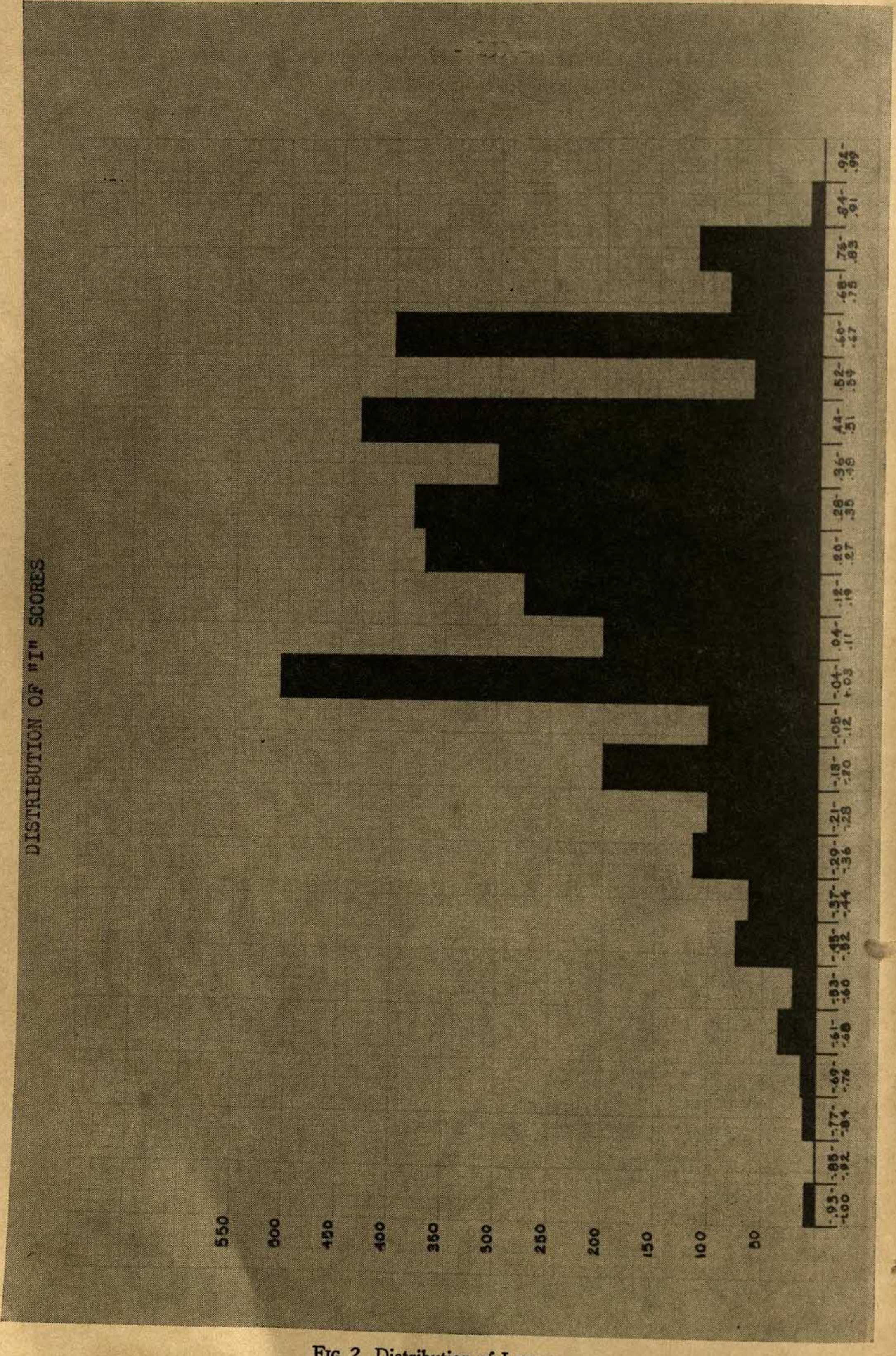


Fig. 2. Distribution of I scores.

TABLE 5

Analysis of Variance Block Scores

D SCORES

Area	Scores	N	X
1	15230	972	15.668
2	11133	972	11.453
3	11928	972	12.271
4	13627	972	14.019
	51918		
[b] Standard error (area) = .614			
Trio			
1	18814	1296	14.516
2	16289	1296	12.568
3	16815	1296	12.974
	51918		
[b] Standard error (trio) = .533			
Triad			
1	3701	324	11.422
2	4267	324	13.169
3	3892	324	12.012
4	4809	324	14.842
5	4985	324	15.386
6	4280	324	13.209
7	3482	324	10.746
8	4166	324	12.858
9	5382	324	16.611
10	4579	324	14.132
11	4928	324	15.209
12	3447	324	10.638
	51918		
[b] Standard error (triad) = 1.063			
Instrument			
1	17195	1296	13.267
2	16664	1296	12.858
3	18059	1296	13.934
	51918		
[b] Standard error (instrument) = .533			
Order			
1	18905	1296	14.587
2	16555	1296	12.773
3	16458	1296	12.699
	51918		
[b] Standard error (order) = .533			
Sequence[a]			
0	17177	1296	13.253
1	5163	420	12.292
2	6734	456	14.767
3	5358	420	12.757
4	5716	384	14.885
5	805	36	22.361
6	4462	396	11.267
7	608	60	10.133
8	5154	372	13.854
9	741	48	15.437
	51918		

Subject	Scores	N	X
1	11888	108	11.000
2	899	108	8.324
3	1614	108	14.944
4	1391	108	12.879
5	1474	108	13.648
6	1402	108	12.981
7	1119	108	10.361
8	1542	108	14.277
9	1231	108	11.398
10	1724	108	15.962
11	1516	108	14.037
12	1569	108	14.527
13	1707	108	15.805
14	1439	108	13.324
15	1839	108	17.027
16	637	108	5.898
17	2027	108	18.768
18	1616	108	14.962
19	1469	108	13.601
20	738	108	6.833
21	1275	108	11.805
22	1505	108	13.935
23	1570	108	14.537
24	1091	108	10.101
25	1935	108	17.916
26	1816	108	16.814
27	1631	108	15.101
28	1628	108	15.074
29	1531	108	14.175
30	1420	108	13.148
31	1561	108	14.453
32	1260	108	11.666
33	2107	108	19.509
34	997	108	9.232
35	1422	108	13.166
36	1028	108	9.518

[b] Standard error (subject) = 1.832

[a] See legend in Sequence table for I scores.

[b] Formula for Standard error $\left(\sqrt{\frac{SS_T - SS_Q}{df_T - df_Q}}\left(\frac{1}{NQ}\right)\right)$ assumes equal variances in all levels of the factor.

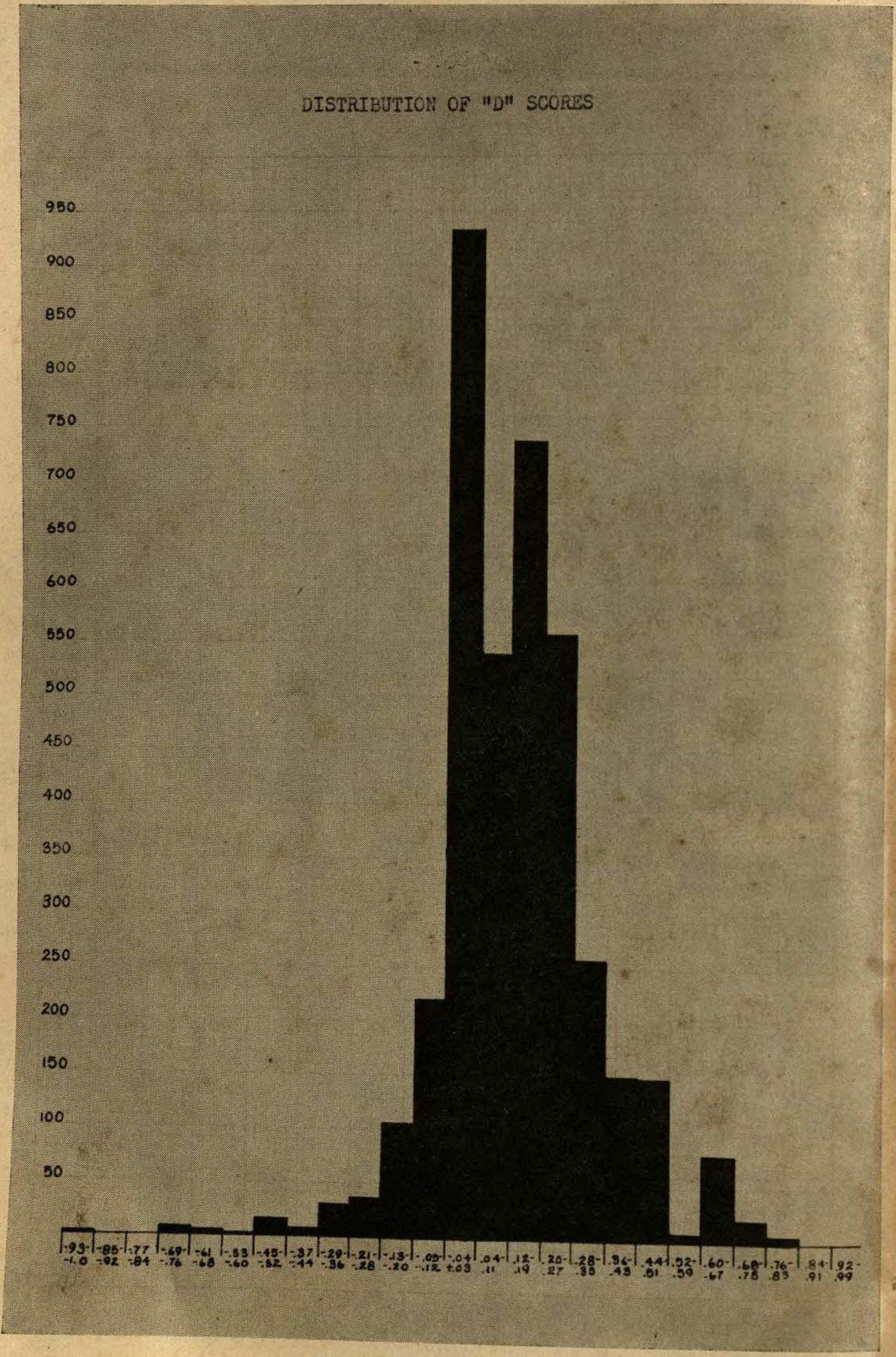


FIG. 3. Distribution of D scores.

TABLE 6

INTERJUDGE AGREEMENT

Trio #1

(Judges 1 and 4)

		Judge 4		
		Like	Unlike	Total
Judge 1	Like	681	372	1053
	Unlike	341	1113	1454
	Total	1022	1485	2507

$\phi = .414$

(Judges 1 and 7)

		Judge 7		
		Like	Unlike	Total
Judge 1	Like	760	297	1057
	Unlike	464	991	1455
	Total	1224	1288	2512

$\phi = .516$

Trio #2

(Judges 2 and 5)

		Judge 5		
		Like	Unlike	Total
Judge 2	Like	774	451	1225
	Unlike	433	842	1275
	Total	1207	1293	2500

$\phi = .380$

Trio #2

(Judges 2 and 8)

		Judge 8		
		Like	Unlike	Total
Judge 2	Like	831	376	1207
	Unlike	277	1005	1282
	Total	1108	1381	2489

$\phi = .475$

Trio #3

(Judges 3 and 6)

		Judge 6		
		Like	Unlike	Total
Judge 3	Like	810	373	1183
	Unlike	284	965	1249
	Total	1094	1338	2432

$\phi = .459$

(Judges 3 and 9)

		Judge 9		
		Like	Unlike	Total
Judge 3	Like	737	451	1188
	Unlike	234	1016	1250
	Total	971	1467	2438

$\phi = .442$

TABLE 7

OCCURRENCE OF ITEMS OF "NO INFORMATION" SUMMARIZED FOR BEHAVIOR AREAS AND CRITERION INSTRUMENTS

Area of Relationships	Statement Items Included	Criterion Instruments Number of Items and Percentages						Area Totals	Preferential Occurrence of "No Information" to Particular Items
		Pencil-and-Paper		Rorschach		Observation and Interview			
Peer group	1-3, 26, 27	*N* 90	35.7[b]	*N* 54	21.4[b]	*N* 108	42.8[b]	252	100% of "No Information" items for this area found in Items 26 and 27
		6.3[a]		6.9[a]		9.5[a]			
Family	4-9	*N* 342	42.3	*N* 404	50.0	*N* 63	7.8	809	
		23.9		51.7		5.6			
Intellect	10-15	*N* 387	46.3	*N* 126	15.1	*N* 324	38.7	837	50% of "No Information" items for this area found in Item 14, 70% found in Items 14 and 15
		27.0		16.1		28.5			
Emotional Adjustment	16-25	*N* 612	42.2	*N* 198	13.7	*N* 639	44.1	1449	75% of "No Information" items fo this area found i Items 19, 22, and 25
		42.7		25.3		56.3			
Totals		1431		782		1134		3347	

[a] Entries in this column indicate percentage of total Instrument items of no information contained in the particular Area-Instrumen block.

[b] Entries in this column indicate percentage of total Area items of no information contained in the particular Area-Instrument bloc

EDUCATIONAL AND PSYCHOLOGICAL
GOVT OF W. BENGAL
BUREAU OF

REHABILITATION-INDIA
CALCUTTA
70300